DOCUMENT RESUME

ED 479 472 UD 035 825

AUTHOR Lutkus, Anthony D.; Weiner, Arlene W.; Daane, Mary C.; Jin, Ying; Jerry, Laura; Kokolis, Gerry

TITLE The Nation's Report Card: Writing 2002, Trial Urban District Assessment.

INSTITUTION National Center for Education Statistics (ED), Washington, DC.; Educational Testing Service, Princeton, NJ.

REPORT NO NCES-2003-530

PUB DATE 2003-07-00

NOTE 81p.; In collaboration with Laura Jerry, Gerry Kokolis. For the writing assessment results, see UD 035 825.

AVAILABLE FROM ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-4ED-PUBS (Toll Free). Fax: 301-470-1244. Web site: http://www.edpubs.org. For full text: http://www.nces.ed.gov/pubsearch.

PUB TYPE Reports - Research (143)

EDRS PRICE EDRS Price MF01/PC04 Plus Postage.

DESCRIPTORS Academic Achievement; Educational Assessment; Educational Status Comparison; Elementary Education; Grade 4; Grade 8; National Surveys; School Districts; *Test Results; *Urban Schools; Writing (Composition); *Writing Achievement; *Writing Tests

IDENTIFIERS Atlanta Public Schools GA; Chicago Public Schools IL; District of Columbia Public Schools; Houston Independent School District TX; Los Angeles Unified School District CA; *National Assessment of Educational Progress; Test Trials

ABSTRACT

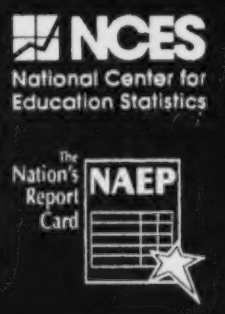
In 2001 Congress appropriated funds for a trial district-level assessment and the National Assessment Governing Board passed a resolution approving the selection of five large urban school districts for participation in the Trial Urban District Assessment, a special project within the National Assessment of Educational Progress (NAEP). This report presents, for the first time, district-level results for NAEP writing assessments in these school districts: Atlanta City, Georgia; Chicago, Illinois; Houston Independent School District, Texas; Los Angeles Unified School District, California; and New York City Public Schools. Results are also included for the District of Columbia. The writing assessments were administered in grades 4 and 8 to representative samples of students. Results for New York City for grade 8 are not reported because they did not meet statistical standards for representation. Average scores for fourth graders in public schools ranged from 135 in the District of Columbia to 153 in New York City and the nation. Average scores for eighth graders ranged from 128 in the District of Columbia and Los Angeles to 138 in Houston. In each of these districts, the average score was lower than the average score for eighth graders in the nation. At grade 4, the percentages of students performing at or above "Proficient" ranged from 11% in the District of Columbia to 27% in New York City. The percentage of students performing at or above "Proficient" was higher for the nation than for any of the five urban districts reported, and higher for central cities than for all the urban districts except Houston, for which no significant difference was detected.

Reproductions supplied by EDRS are the best that can be made from the original document.

Results are also given for student subgroups by gender, race/ethnicity, eligibility for free/reduced price lunch, and parents' highest level of education. Three appendixes provide information about study procedures and methods of analysis. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 479 472



U.S. Department of Education
Institute of Education Sciences
NCES 2003-530

The Nation's Report Card

Trial Urban District Assessment

Writing 2002

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

The National Assessment of Educational Progress

UD 035 825

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is a nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations.

In 1988, Congress established the National Assessment Governing Board (NAGB) to oversee and set policy for NAEP. The Board is responsible for: selecting the subject areas to be assessed; setting appropriate student achievement levels; developing assessment objectives and test specifications; developing a process for the review of the assessment; designing the assessment methodology; developing guidelines for reporting and disseminating NAEP results; developing standards and procedures for interstate, regional, and national comparisons; determining the appropriateness of all assessment items and ensuring the assessment items are free from bias and are secular, neutral, and non-ideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of National Assessment of Educational Progress reports.

The National Assessment Governing Board

Darvin M. Winick, Chair
President
Winick & Associates
Dickinson, Texas

Amanda P. Avallone
Assistant Principal and
Eighth-Grade Teacher
Summit Middle School
Boulder, Colorado

Daniel A. Domenech
Superintendent of Schools
Fairfax County Public Schools
Fairfax, Virginia

Edward Donley
Former Chairman
Air Products & Chemicals, Inc.
Allentown, Pennsylvania

Honorable Dwight Evans
State Legislator
Philadelphia, Pennsylvania

Thomas H. Fisher
Director (Retired)
Student Assessment Services
Florida Department of Education
Tallahassee, Florida

Sheila M. Ford
Principal
Horace Mann Elementary School
Washington, DC

Edward H. Haertel
Professor, School of Education
Stanford University
Stanford, California

Catherine Harvey
Principal
Bethesda-Chevy Chase High School
Bethesda, Maryland

Juanita Haugen
Local School Board Member
Pleasanton, California

Honorable Dirk Kempthorne
Governor of Idaho
Boise, Idaho

Kim Kozbial-Hess
Fourth-Grade Teacher
Fall-Meyer Elementary School
Toledo, Ohio

Honorable Ronnie Musgrove
Governor of Mississippi
Jackson, Mississippi

Mark D. Musick
President
Southern Regional Education Board
Atlanta, Georgia

Honorable Jo Ann Pottorff
State Legislator
Wichita, Kansas

Diane Ravitch
Senior Research Scholar
New York University
New York, New York

Sister Lourdes Sheehan, R.S.M.
Associate General Secretary
United States Catholic Conference
Washington, DC

Honorable Raymond Simon
Director
Arkansas Department of Education
Little Rock, Arkansas

John H. Stevens
Executive Director
Texas Business and Education Coalition
Austin, Texas

Deborah Voltz
Associate Professor
Department of Special Education
University of Louisville
Louisville, Kentucky

Honorable Michael E. Ward
State Superintendent of Public Instruction
Public Schools of North Carolina
Raleigh, North Carolina

Marilyn A. Whirry
Twelfth-Grade English Teacher
Manhattan Beach, California

Dennie Palmer Wolf
Director of Opportunity
and Accountability
Annenberg Institute for School Reform
Brown University
Providence, Rhode Island

Honorable Grover (Russ) Whitehurst (Ex-Officio)
Director
Institute of Education Sciences
U.S. Department of Education
Washington, DC

Charles E. Smith
Executive Director, NAGB
Washington, DC

BEST COPY AVAILABLE

NCES
National Center for Education Statistics

The Nation's Report Card
Writing 2002
TRIAL URBAN DISTRICT ASSESSMENT

U.S. Department of Education
Institute of Education Sciences
NCES 2003-530

Anthony D. Lutkus
Mary C. Daane
Arlene W. Weiner
Ying Jin

in collaboration with

Laura Jerry
Gerry Kokolis
Educational Testing Service

Taslima Rahman
Project Officer
National Center for Education Statistics

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Val Plisko
Associate Commissioner

July 2003

SUGGESTED CITATION
U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. *The Nation's Report Card: Writing 2002, Trial Urban District Assessment,* NCES 2003-530, by A. D. Lutkus, M. C. Daane, A. W. Weiner, and Y. Jin. Washington, DC: 2003.

FOR MORE INFORMATION
Content contact:
Taslima Rahman
202-502-7316

To obtain single copies of this report, or ordering information on other U.S. Department of Education products, call toll free 1-877-4ED-PUBS (877-433-7827), or write:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794-1398

TTY/TDD 1-877-576-7734
FAX 301-470-1244

Online ordering via the Internet: *http://www.edpubs.org*
Copies also are available in alternate formats upon request.
This report also is available on the World Wide Web: *http://nces.ed.gov/pubsearch*

The work upon which this publication is based was performed for the National Center for Education Statistics by Educational Testing Service.

Table of Contents

Executive Summary . . . . . **vii**

Chapter 1
Introduction **1**

Overview of the Trial Urban District Assessment in Writing 2002 1
Brief History of the National Assessment of Educational Progress 1
Background of the NAEP Trial Urban District Assessment in 2002 2
Selection of Urban Districts 3
Overview of the NAEP 2002 Writing Assessment 3
Objectives and Content of the Assessment 3
The 2002 Writing Assessment Procedure 6
Procedures for Sampling Student Populations 7
Evaluating Students' Writing 9
Reporting the Writing Assessment Results 9
Interpreting NAEP Results 11
Cautions in Interpretations 12

Chapter 2
Average Scale Score and Achievement Level Results for the Trial Urban District Assessment **15**

Urban District Scale Score and Percentile Results 15
Comparisons Among Districts by Average Scale Scores 18
Writing Achievement Level Results 21
Comparisons Among Districts by Achievement Levels 22

Chapter 3
Results for Subgroups **25**

Performance of Selected Subgroups in the Trial Urban District Assessment 26
Gender 26
Race/Ethnicity 30
Free/Reduced-Price Lunch Eligibility 37
Parents' Highest Level of Education 40

Appendix A
Identification, Exclusion, and Accommodation Rates for Special-Needs Students **43**

Appendix B
District-Level Contextual Variables **51**

Appendix C
Overview of Procedures Used for the NAEP 2002 Writing Trial Urban District Assessment **59**

Acknowledgments **81**

Chapter 1: Tables and Figures

Figure 1.1
Descriptions of the three purposes for writing in the NAEP writing assessment 4

Table 1.1
Distribution of writing prompts, by purpose for writing, in the NAEP 2002 writing assessment, grades 4 and 8 5

Figure 1.2
Descriptions of NAEP writing achievement levels, grade 4 13

Figure 1.3
Descriptions of NAEP writing achievement levels, grade 8 14

Chapter 2: Tables and Figures

Table 2.1
Average writing scale scores and selected percentiles, grades 4 and 8 public schools: By urban district, 2002 17

Figure 2.1
Cross-district comparisons of average writing scale scores, grade 4 public schools: By urban district, 2002 19

Figure 2.2
Cross-district comparisons of average writing scale scores, grade 8 public schools: By urban district, 2002 20

Table 2.2
Percentage of students at or above each writing achievement level, grades 4 and 8 public schools: By urban district, 2002 21

Figure 2.3
Cross-district comparisons of percentage of students at or above *Proficient* in writing, grade 4 public schools: By urban district, 2002 23

Figure 2.4
Cross-district comparisons of percentage of students at or above *Proficient* in writing, grade 8 public schools: By urban district, 2002 24

Chapter 3: Tables and Figures

Table 3.1
Average writing scale scores and percentage of students at or above each achievement level, by gender, grade 4 public schools: By urban district, 2002 27

Table 3.2
Average writing scale scores and percentage of students at or above each achievement level, by gender, grade 8 public schools: By urban district, 2002 28

Figure 3.1
Gaps in average writing scale scores, by gender, grades 4 and 8 public schools: By urban district, 2002 29

Table 3.3
Average writing scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 4 public schools: By urban district, 2002 32

Table 3.4
Average writing scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 8 public schools: By urban district, 2002 34

Figure 3.2
Gaps in average writing scale scores, by race/ethnicity, grades 4 and 8 public schools: By urban district, 2002 36

Table 3.5
Average writing scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 4 public schools: By urban district, 2002 38

Table 3.6
Average writing scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 8 public schools: By urban district, 2002 39

Table 3.7
Average writing scale scores and percentage of students at or above each achievement level, by student-reported parents' highest level of education, grade 8 public schools: By urban district, 2002 41

Executive Summary

The National Assessment of Educational Progress (NAEP) is the nation's ongoing representative sample survey of student achievement in core subject areas. NAEP, known as the Nation's Report Card, is authorized by Congress and administered by the National Center for Education Statistics (NCES) of the Institute of Education Sciences in the U.S. Department of Education. NAEP regularly reports to the public on the educational progress of students in grades 4, 8, and 12.

In 2002, NAEP assessed the reading and writing performance of the nation's fourth-, eighth-, and twelfth-grade students. NAEP also conducted assessments of fourth- and eighth-graders' reading and writing in most of the states.

In 2001, after discussion among NCES, the National Assessment Governing Board (NAGB), and the leadership of the Council of the Great City Schools, Congress appropriated funds for a trial district-level assessment and NAGB passed a resolution approving the selection of five large urban districts for participation in the Trial Urban District Assessment, a special project within NAEP. This report presents results of NAEP's Trial Urban District Assessment in writing for public school students in the following participating urban school districts: Atlanta City, Chicago School District 299, Houston Independent School District, Los Angeles Unified, New York City Public Schools, and Washington, DC. This represents NAEP's first assessment of urban districts based on samples specially

designed to allow reporting of subgroup data. The first five districts participated voluntarily in the NAEP 2002 writing assessment at grades 4 and 8. Results for the District of Columbia, which in this and past NAEP assessments has been sampled and assessed along with states and other jurisdictions, are also included in this report. Data for public schools across the nation and for central city public schools

are provided for comparison purposes.[1] The public schools sampled also included charter schools, which in some cases were not managed by the urban school districts.

NAEP does not provide scores for individual students or schools. It reports results for groups of students (e.g., fourth-graders). For each group on each table in the report, assessment results are described in one of two ways. First, the group's average writing score is reported on a scale from 0 to 300. Performance for each grade is scaled separately; therefore, average scale scores cannot be compared across grades. The term "average score" is used throughout this report to refer to the average scale score on the NAEP writing scale. Second, student writing performance is reported in terms of the percentage of students in the group who reached each of three achievement levels: *Basic*, *Proficient*, and *Advanced*. The *Proficient* level for each grade is defined by NAGB as representing "solid academic performance," which demonstrates "competency over challenging subject matter" for the grade assessed. *Basic* indicates partial mastery of skills that are fundamental for proficient work. *Advanced* denotes superior performance.

The achievement levels are performance standards adopted by the National Assessment Governing Board (NAGB) as part of its statutory responsibilities. The achievement levels are a collective judgment of what students should know and be able to do for each grade tested. As provided by law, NCES, upon review of a congressionally mandated evaluation of NAEP, determined that the achievement levels are to be used on a trial basis and should be interpreted with caution. However, both NCES and the Board believe that the performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials and others as a common yardstick of academic performance.

The results are based on representative samples of students for the nation, for participating districts, and for students in schools in central cities. A central city is a city of 50,000 or more that is the largest in its metropolitan area, or can otherwise be regarded as "central." The term means "a city that is central," not "the central part of a city" or the "inner city." Note that central cities encompass wider areas than what is commonly referred to as "the inner city." (See further details in appendix C, Type of Location).

In order to obtain reliable data, sufficient numbers of the selected schools and students must participate in the assessment. All six districts met the NCES participation criteria for NAEP samples at grade 4, but results for New York City schools at grade 8 are not reported because they did not meet the participation criteria.

[1] "Central city" is defined in chapter 2 and more completely in the "Type of Location" section of appendix C. Central city includes nationally representative public schools located in central cities within metropolitan statistical areas as defined by the Federal Office of Management and Budget. It is not synonymous with "inner city."

Some students are identified by the school districts as students with disabilities and/or limited English proficient students. Some of these students are excluded from the assessment, and others are tested with accommodations related to their status. Three of the six districts identified between 30 and 52 percent of their students as either students with disabilities or limited English proficient students. Because the percentage of students identified, excluded, and assessed with accommodations varies across the districts, that variability should be taken into consideration in interpreting the results and making comparisons (See appendix A, table A.1).

Throughout this report, differences between scores and between percentages are discussed in the text in terms of statistical significance. All differences reported are significant at the 0.05 level (with appropriate adjustments for multiple comparisons).

Overall Writing Results for the Urban Districts

The following summary first describes results in terms of average scale scores and then in terms of achievement levels. Average results for public schools in the districts participating in the Trial Urban District Assessment are compared, at grades 4 and 8, with public schools in the nation, with public schools in central cities, and with each other.

Average Scale Scores

Results for Grade 4 Public Schools

- The average scores for fourth-graders in public schools ranged from 135 in the District of Columbia to 153 in New York City and the nation.
- At grade 4, no statistically significant differences were detected between the average scores for students in Houston and New York City and the average score for students in the nation, while students in Atlanta, Chicago, the District of Columbia, and Los Angeles had average scores lower than the average score in the nation.
- At grade 4, the average score for students in New York City was higher than the national average score for students in central city schools. The average score for fourth-graders in Houston was not found to differ significantly from that for central cities, while the average score for students in each of the other districts was lower than the national average score for central cities.
- The average score for students in New York City was higher than those in all the other participating districts except Houston. The average score in Houston was higher than the scores in Atlanta, Chicago, and the District of Columbia, but was not found to differ significantly from the average scores in Los Angeles and New York City.

Results for Grade 8 Public Schools

Results for New York City schools at grade 8 are not reported because they did not meet participation criteria.

- The average district scores for eighth-graders ranged from 128 in the District of Columbia and Los Angeles to 138 in Houston.
- In each of the reported districts, the average score of eighth-grade students was lower than the average score for eighth-grade students in the nation.

- At grade 8, no significant difference was detected between the average score for students in Houston and the average score for students in the central city schools. The average score in the central city schools was higher than the average scores in Atlanta, Chicago, the District of Columbia, and Los Angeles.
- The average score for students in Houston was higher than the average scores in Atlanta, the District of Columbia, and Los Angeles. The average score in Chicago was not found to differ significantly from those in Houston and Atlanta, and was higher than the average scores in the District of Columbia and Los Angeles.

Writing Achievement Levels

Results for Grade 4 Public Schools

- At grade 4, the percentages of students performing at or above *Proficient* ranged from 11 percent in the District of Columbia to 27 percent in New York City.
- The percentages of fourth-grade students performing at or above *Proficient* in Houston and New York City were not found to be significantly different from the percentages in the nation or in central cities. Atlanta, Chicago, the District of Columbia, and Los Angeles had lower percentages of students at or above *Proficient* than the nation and central cities.
- At grade 4, the percentage of students performing at or above *Proficient* in New York City was higher than the percentages in four other districts and not found to differ significantly from the percentage in Houston.
- The percentages of fourth-grade students performing at or above *Basic* ranged from 73 percent in the District of Columbia to 85 percent in New York City. In public schools across the nation, 85 percent of students performed at or above the *Basic* level. In central city schools, 81 percent performed at or above the *Basic* level.

Results for Grade 8 Public Schools

- At grade 8, the percentages of students performing at or above *Proficient* ranged from 10 percent in Atlanta and the District of Columbia to 19 percent in Houston. Thirty percent of eighth-graders in public schools in the nation and 22 percent in central city schools performed at or above the *Proficient* level.
- The percentage of students performing at or above *Proficient* was higher for the nation than for any of the five urban districts reported, and higher for central cities than for all urban districts except Houston, where no significant difference was detected. The percentages of eighth-graders performing at or above *Proficient* in Chicago and Houston were not found to differ significantly from each other, and both were higher than the comparable percentages in the District of Columbia and Atlanta.
- The percentage of eighth-graders performing at or above *Basic* ranged from 64 percent in Los Angeles to 74 percent in Houston. Eighty-four percent of eighth-graders in public schools in the nation and 77 percent in central city public schools performed at or above the *Basic* level.

Results for Student Subgroups

In addition to providing average scores and achievement levels for the nation, for states, and, in this report, for districts, NAEP reports provide results for subgroups of students defined by various background and contextual characteristics (e.g., gender, eligibility for free/reduced-price lunch, and level of parents' education). Performance results for subgroups are reported primarily as comparisons of district average scores with the comparable average scores in central cities.

Gender

Results for Grade 4 Public Schools

- No statistically significant difference was detected between the average scores of male or female fourth-grade students in Houston, New York City and the average scores of their counterparts in the central city schools. Average scores for fourth-grade male and female students in Atlanta, Chicago, the District of Columbia, and Los Angeles were lower than the average score in central city schools.
- Female fourth-graders had higher average scores than male fourth-graders in each of the urban districts.

Results for Grade 8 Public Schools

- The average score for eighth-grade female students in Houston was not found to be significantly different from that of the central city public schools. In Atlanta, Chicago, the District of Columbia, and Los Angeles, the average scores for both male and female eighth-graders were lower than the average scores for their counterparts in central city schools.
- In all participating districts, female students had higher average writing scores than male students.

Race/Ethnicity

In each of the urban districts assessed, Black or Hispanic students constitute the majority or the largest racial/ethnic group. This distribution differs from that for the national writing assessment, in which White students constitute a majority—60 percent of the fourth-grade sample and 64 percent of the eighth-grade sample. Black students made up more than four-fifths of the samples at both grades in Atlanta and the District of Columbia and nearly half at both grades in Chicago. Hispanic students made up about two-thirds of the Los Angeles samples at both grades and about half of the fourth-graders and more than half of the eighth-graders in Houston. In New York City, more than two-fifths of the fourth-graders were Hispanic and just under a third were Black.

Results for Grade 4 Public Schools

- In five of the six urban districts in which a reliable comparison could be made, White fourth-graders had higher average scores than their Black and Hispanic counterparts.
- Black students in grade 4 in Houston and New York City had higher average scores than those in the central cities. Black fourth-grade students in Atlanta, Chicago, and Los Angeles had average scores not found to differ significantly from their counterparts in central cities. In the District of Columbia, Black fourth-graders had an average score lower than that of their counterparts in central cities.
- No significant difference was detected between the average score for Hispanic fourth-graders in five of the six districts and Hispanic fourth-graders' average score in central cities. The average score for Hispanic fourth-graders in Los Angeles was lower than that in central cities taken as a whole.

- Average scores for White fourth-grade students in Atlanta, the District of Columbia, and New York City were higher than the average score for White fourth-grade students in central cities.
- Average scores for Asian/Pacific Islander students in Los Angeles and New York City were not found to be significantly different from the average score for their counterparts in central cities.

Results for Grade 8 Public Schools

- White eighth-graders had higher average scores than Black eighth-graders, in every district except the District of Columbia, where the sample size was insufficient to permit a reliable comparison. White students at grade 8 also had higher average scores than Hispanic students in Chicago, Houston, and Los Angeles.
- The average score for eighth-grade Black students in the District of Columbia was lower than that for Black eighth-grade students in the central city schools, and no significant difference was detected between the average score for Black students in any of the other four districts and the national average score for Black students in central cities.
- Average scores for Hispanic students were not found to differ significantly between the districts and the national average for central cities, except in Los Angeles, where Hispanic students had a lower average score than their counterparts in central cities.
- The average score for White eighth-grade students in Houston was higher than that of White students in the central city schools, while the average score in Los Angeles was lower.

Eligibility for Free/Reduced-Price Lunch

The National School Lunch Program providing free/reduced-price lunch is administered by the U.S. Department of Agriculture (USDA) for children near or below the poverty line. Eligibility is determined by the USDA's Income Eligibility Guidelines (http://www.fns.usda.gov/cnd/IEGs&NAPs/IEGs.htm).

Results for Grade 4 Public Schools

At grade 4, rates of student eligibility for free/reduced-price lunch ranged from 70 percent in New York City to 89 percent in Chicago.

- Fourth-grade students eligible for free/reduced-price lunch had lower average scores than those not eligible in every district except Los Angeles and New York City, where no significant difference between the two eligibility categories was detected.
- Fourth-grade students eligible for free/reduced-price lunch in New York City had a higher average score than the national average score for their counterparts in central cities, while students in Chicago and the District of Columbia had lower average scores than their eligible counterparts in central city schools.
- The average scale score for ineligible students at grade 4 in the District of Columbia was lower than the national average score in central city schools.

Results for Grade 8 Public Schools

At grade 8, the percentages of eligible students ranged from 67 percent in the District of Columbia to 84 percent in Chicago. Because the available data for eligibility for eighth-graders in Los Angeles did not meet reporting standards, no information related to eligibility is reported for this segment of the sample.

- Students at grade 8 who were not eligible for free/reduced-price lunch had a higher average score than eligible students in every district where the data were sufficiently reliable for significance testing.
- At grade 8, both those students eligible for free/reduced-price lunch and those not eligible in Atlanta and the District of Columbia had lower average scores than their counterparts in the central city public schools.

Parents' Highest Level of Education

Eighth-grade students who participated in the Trial Urban District Assessment were asked to indicate the highest level of education their parents had completed. Five response options—did not finish high school, graduated from high school, some education after high school, graduated from college, or "I don't know"—were offered.

- In all five districts, lower percentages of students reported that their parents had graduated from college than in the national sample. The percentages of students who reported that their parents did not graduate from high school were higher than the nation in Chicago, Houston, and Los Angeles.
- Atlanta and the District of Columbia had the highest percentages of students who reported that at least one parent had graduated from college (35 and 37 percent, respectively). These percentages were significantly higher than those in Houston and Los Angeles. Atlanta also had a higher percentage of students reporting parents with some education after high school than all the other districts. Houston and Los Angeles had the highest percentages of students reporting parents who did not finish high school (22 and 18 percent, respectively).
- Average scores in all districts except Houston were lower for students who reported a college graduate parent than the national average score for their counterparts in the central city public schools.
- In Chicago and Houston, no statistically significant difference was detected between the average score of students with parents who did not finish high school and the average score of their counterparts in the central city schools, while the average score of these students in the Atlanta, the District of Columbia, and Los Angeles was lower than the national average score in the central city schools.

1 Introduction

Overview of the Trial Urban District Assessment in Writing 2002

Writing is an important tool for communicating ideas in school, work, and the community. Therefore, writing has always been central to elementary and secondary school curricula. The National Assessment of Educational Progress (NAEP) 2002 Writing Trial Urban District Assessment (TUDA) initiates a national endeavor to provide the public with reliable evidence about urban fourth- and eighth-grade students' achievements in writing, and makes comparisons to public schools nationally and central city schools.[1]

Brief History of the National Assessment of Educational Progress

For more than thirty years, NAEP has been providing reliable information about American students' achievement. The purpose of NAEP reports is to inform educators, policy makers, parents, and the public. In order to do so, NAEP (also known as the Nation's Report Card) regularly and systematically collects, analyzes, and reports valid and reliable information about what American students know and can do in a variety of subject areas. NAEP assesses representative national samples of fourth-, eighth-, and twelfth-grade students, and representative samples of students in a number of states and other jurisdictions, including the District of Columbia and the Department of Defense schools (domestic and overseas).

[1] "Central city" is defined in chapter 2 and more completely in the "Type of Location" section of appendix C. Central city includes nationally representative public schools located in central cities within metropolitan statistical areas as defined by the Federal Office of Management and Budget. It is not synonymous with "inner city."

Authorized by Congress in 1969, NAEP is a continuing expression of the nation's concern with students' learning. The National Center for Education Statistics (NCES), one of three centers within the U.S. Department of Education's Institute of Education Sciences (IES), manages NAEP. The National Assessment Governing Board (NAGB), an independent body, provides policy guidance for NAEP.

NAEP does not report results for individual students or schools. In fact, for its first two decades, NAEP reported results only for the nation as a whole and for subgroups within the nation (e.g., for female students and male students).

In 1988, Congress authorized a trial of state-level assessments. In 1990, NAEP undertook its first state-level assessment, of the mathematics performance of a representative sample of students in participating states. This successful trial of state-level assessments led to a broadening of the endeavor, and by 1994, NAEP was reporting to many states on their students' performance in mathematics, reading, writing, and science. In 2002, NAEP assessed representative samples of students in the nation, in most of the states, and also in several U.S. jurisdictions. In addition, in 2002 NAEP assessed the performance of public-school students in a set of school districts. This report is the initial presentation of writing assessment results for each of five participating urban school districts and the District of Columbia.

The national-, state-, and district-level reports of the NAEP 2002 assessment are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/writing/).

The national- and state-level reports present nationwide results for grades 4, 8, and 12, and results for grades 4 and 8 within the states and other jurisdictions that participated in the state-level assessment. The national and state reports also compare the results to the results of previous NAEP reading assessments.

Background of the NAEP Trial Urban District Assessment in 2002

Over the years, various constituents have requested reports on individual school results and district results, but the NAEP samples have not been designed to permit reporting of reliable data below the state level. The District of Columbia is an exception. The NAEP legislation (as well as other federal education programs) defines the District of Columbia as a state and authorizes its participation in state NAEP. Through its participation in state NAEP, student achievement data for the District of Columbia school system are regularly reported as part of the NAEP program.

Federal appropriations authorized for the No Child Left Behind Act supported a multi-year study of the feasibility of a Trial Urban District Assessment as a component of the National Assessment of Educational Progress.[2] Some large urban school districts, such as Los Angeles and New York, have enough students to meet NAEP requirements for sample size in reporting. In 2001, after discussion among the National Assessment Governing Board, the National Center for Educational Statistics, and the leadership of the Council of the Great City Schools, NAGB passed a resolution approving the selection of five large urban districts for participation in a

[2] No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Trial Urban District Assessment, a special project within NAEP.[3]

The trial assessment called for larger-than-usual samples within the districts, making reliable district-level data possible. The assessment allows these districts to make valid comparisons among themselves for the first time. Although individual states' own assessments may report data by district and school, their results are based on different scores, scales, and test designs; therefore, until now, districts have not been able to compare themselves to districts in other states.

By undertaking the Trial Urban District Assessment in reading and writing in 2002, NAEP continues a tradition of carefully extending its service to education, while preserving the rigorous sampling, scoring, and reporting procedures that have characterized the national and state assessments. The samples were large enough to provide data on subgroups within the districts, such as female students or Hispanic students. Thus, the 2002 data can serve the districts as a benchmark for studying changes in the performance of all their students and of particular subgroups of students.

In addition to assessing subject-area performance of groups and subgroups, NAEP gathers contextual data about in- and out-of-school experiences and socioeconomic factors from background questionnaires given to students, teachers, and school administrators. Large amounts of additional data not included in this report are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

Selection of Urban Districts

NAGB staff, assisted by representatives of the Council of the Great City Schools, identified five districts for the trial assessment. Districts were selected that permitted testing the feasibility of conducting NAEP over a range of characteristics, such as district size, minority concentrations, federal program participation, poverty, and percentages of students with disabilities and limited English proficient students. This report presents data for the following participating urban districts: Atlanta City, Chicago School District 299, Houston Independent School District, Los Angeles Unified, New York City Public Schools, and Washington, DC. All these participating districts are located in central cities.

Overview of the NAEP 2002 Writing Assessment

Objectives and Content of the Assessment

Each NAEP assessment has objectives described in a "framework"—a document that specifies the important content and process areas to be measured and the types of questions to be included in the assessment. NAGB directs a process for specifying these frameworks.

The NAEP 1998 writing framework is the blueprint that has specified the content and guided the development of the 1998 and 2002 writing assessments.[4] The framework establishes the assessment objectives and provides direction on the kinds of writing tasks to be included in the instrument. A task, or writing assessment item, is usually a short text or stimulus

[3] National Assessment Governing Board. (2001, November 20). Minutes of the November 15–17, 2001 NAGB Meeting Committee Reports and Board Actions. Presented at the November 2001 NAGB meeting, Washington, DC.

[4] National Assessment Governing Board. (1998). *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: Author.

posing a situation, concern, or topic about which students are asked to write under a stated time constraint.

The framework resulted from a national process involving many parties concerned about writing education, including teachers, state education officials, subject-area specialists, researchers, and representatives of the general public. This effort was managed by the Center for Research on Evaluation, Standards, and Student Testing (CRESST), under the direction of NAGB. NAGB also contracted with ACT to develop detailed specifications for assessment items.

The NAEP writing framework, informed by current research and theory, emphasizes that writing addresses a variety of purposes and audiences. The framework discusses three purposes for writing: **narrative** (telling a story), **informative** (informing the reader), and **persuasive** (persuading the reader). To ensure that NAEP writing assessments reflect the genres receiving the most instructional emphasis, the framework prescribes that NAEP writing tasks focus on these three purposes at grades 4, 8, and 12.[5] Descriptions of narrative, informative, and persuasive writing appear in figure 1.1.

Figure 1.1 Descriptions of the three purposes for writing in the NAEP writing assessment

Purposes for Writing	
Narrative writing	Narrative writing encourages writers to incorporate their imagination and creativity in the production of stories or personal essays. At its best, narrative writing fosters imagination, creativity, and speculation by allowing writers to express their thoughts and emotions, and offers an opportunity for writers to analyze and understand their actions and those of others. The narrative prompts included in the NAEP 2002 writing assessment asked students to write many kinds of stories (most fiction, some nonfiction). Some of the prompts asked students to write in response to photographs, drawings, cartoons, poems, or stories (provided with the assessment).
Informative writing	In informative writing, the writer provides the reader with information. Informative writing may involve reporting on events or experiences or analyzing concepts and relationships. When used as a means of exploration, informative writing helps both the writer and the reader to learn new ideas and to reexamine old conclusions. Informative prompts in the NAEP 2002 writing assessment asked students to write on specified subjects using many kinds of information, such as newspaper articles, charts, photographs, or reported dialogues (provided with the assessment), as well as their own knowledge. Students could write in a variety of formats, such as reports, newpaper articles, and letters.
Persuasive writing	Persuasive writing seeks to persuade the reader to take action or to bring about change. This type of writing involves a clear awareness of what arguments might most affect the audience being addressed. Writing persuasively also requires the use of such skills as analysis, inference, synthesis, and evaluation. Persuasive prompts in the NAEP 2002 writing assessment asked students to write letters to the editor or to friends, to refute arguments, or to take sides in a debate.

SOURCE: National Assessment Governing Board. *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: Author.

[5] On the importance of specifying purpose in writing instruction, see Oliver, E. (1989). Effects of Assignment on Writing Quality at Four Grade Levels. *English Quarterly 21*(4), 224–32.

Gorman, S. (1995). *Windows into the Classroom: NAEP's 1992 Portfolio Study*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

National Assessment Governing Board (1998). *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress*. Washington, DC: Author.

As the framework notes, the purposes for writing are not always completely discrete. For example, a narrative essay may make a persuasive moral or ethical point, and a letter to an editor or congressional representative may include pertinent facts and information. In fact, many of the students whose writing received high ratings used integrated forms of presentation. The professional raters who evaluated the student responses were instructed not to penalize such blended presentations.

The emphasis on each purpose for writing varies from grade to grade to match the differing levels of student development and instructional focus. The assessment emphasized narrative writing for fourth-graders, gave comparable weight to all three purposes for eighth-graders, and stressed persuasive writing for twelfth-graders. Table 1.1 shows both the percentage and actual number of writing tasks for each writing purpose at each grade level in the 2002 assessment. These distributions match the target percentages established by the framework.

Table 1.1 Distribution of writing tasks, by purpose for writing, in the NAEP 2002 writing assessment, grades 4 and 8

	Purpose for writing	Percentage of writing tasks from framework	Actual number of writing tasks
Grade 4			
	Narrative	40	8
	Informative	35	7
	Persuasive	25	5
Grade 8			
	Narrative	35	7
	Informative	35	7
	Persuasive	30	6

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Writing Assessment.

In addition to specifying the percentage of the assessment that should be devoted to each writing purpose, the framework qualifies several elements of writing that should pertain to writing tasks across the assessment. These elements are considered by test developers in the construction of assessment items and are important for motivating student engagement in the assessment tasks. The 2002 assessment used many writing tasks that specified the writer's audience by asking students to write, for example, a letter to a friend or to a school board. Students also had opportunities to write in a variety of forms, such as essays, letters, reports, and stories. Writing tasks may have used any of a variety of stimuli including photographs, cartoons, drawings, newspaper articles, letters, or literary works, such as poems or stories, to evoke written responses. While the limits of a timed assessment prevented students from engaging in the kinds of drafting and revising that can be done in a regular classroom environment, each writing task in the assessment included a planning page to encourage students in that phase of the writing process. In addition, every student received a brief brochure with suggestions for planning and revising their writing. To meet the framework's stated objective that students value writing as a communicative activity, background questions on the assessment asked students about their view of themselves as writers and their writing practices at home and at school.

The 2002 Writing Assessment Procedure

As the only federally authorized, ongoing nationwide assessment of student writing achievement, NAEP writing assessments must reflect the NAEP writing framework and expert perspectives on the measurement of writing competence. To that end, during the development process, the assessment undergoes stringent review by teachers, teacher educators, state officials, and measurement specialists. All components of the assessment are evaluated for curricular relevance, developmental appropriateness, fairness concerns, and adherence to the framework and test specifications. The 2002 writing assessment included twenty writing tasks calling for 25-minute responses at grades 4 and 8. NAEP writing tasks that have been released to the public, along with student performance data by state, are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/itmrls/).

To minimize the burden on any one student, NAEP uses a procedure referred to as "matrix sampling" in which an individual student at any given grade is administered only a small portion of the entire assessment. For example, at each grade, students received test booklets with two writing tasks and two sections of questions regarding their home and school experiences related to writing achievement. Because a representative sample of students at each grade receives each assessment item, the results can then be combined to produce average group and subgroup results based on the entire assessment. In total, the time required for each student to participate in the 2002 NAEP writing assessment was no more than 1 hour.

BEST COPY AVAILABLE

Procedures for Sampling Student Populations

The NAEP 2002 writing assessment was administered to fourth-, eighth-, and twelfth-graders at the national level and to fourth- and eighth-graders at the state and district levels. NAEP uses a random selection process in order to obtain a representative sample of students for reporting national and state or jurisdiction results. Approximately 139,200 fourth-grade students in 5,500 schools; 118,500 eighth-graders in 4,700 schools; and 18,500 twelfth-graders in 700 schools were sampled and assessed. The public schools sampled also included charter schools, which in some cases were not managed by the urban school administration.

The samples were selected using a two-stage sample design that selected schools within participating districts and then selected students within schools. In this report, all references to samples—national, central city, or district—include public-school populations only.

Sampling for the Trial Urban District Assessment was modeled on the procedure for sampling states. The number of participating schools ranged from 38 to 76 per district in the fourth grade and from 15 to 69 per district at the eighth grade. The number of participating students per district ranged from 924 to 2,037 at the fourth grade and from 1,109 to 1,778 at the eighth grade. For information on sample sizes and participation rates by district, see "Urban District Samples" in appendix C.

The overall participation rates for schools and students in the national, state, and urban district assessments must meet statistical guidelines established by NCES and NAGB in order for assessment results to be reported publicly. Data are not reported to the public for a state, jurisdiction, or urban district that participates but does not meet minimum participation rate guidelines. For more information about participation guidelines, see appendix C. Participation rates for the jurisdictions and urban districts were calculated the same way as rates were computed for the nation.

NAEP endeavors to assess all students selected in the random sampling process, including students with disabilities (SD) and/or students who are classified by their schools as limited English proficient (LEP). The percentages of students classified as SD and/or LEP in all participating states and jurisdictions are available in an interactive database at the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/). Information on SD and LEP students for the urban districts appears in appendix A. Percentages of students identified as limited English proficient, particularly at grade 4, appear much higher in some districts (Houston and Los Angeles) than in the nation.

It is important to note that, guided by the student's Individualized Education Program (IEP), as well as eligibility for "section 504" services, school personnel make decisions as to whether students with disabilities should be included in the assessment.[6] Results presented in this report include the performance of students with disabilities who can be meaningfully assessed without accommodations or with the kinds of accommodations NAEP provides, such as small-group settings, extended time, or the use of a scribe or

[6] Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

computer for test administration. Schools also decide whether to include LEP students, based on NAEP's guidelines. The guidelines ask them to judge the student's ability to participate in the assessment in English and to consider the number of years the student has been receiving instruction in English. More information about types of testing accommodations is available in appendix A of this report.

Percentages of students excluded from NAEP may vary considerably across states/districts and, within a state or district, from one year to another. For example, the percentage of fourth-grade students identified as having disabilities and or limited English proficiency ranged from 19 to 30 percent in public schools in the nation, in central cities, and in three of six districts. Atlanta was below the national average, with only 7 percent of its fourth-grade students identified in these groups, and Houston and Los Angeles were above the national average, with 40 and 52 percent, respectively, of their fourth-grade students identified in these groups (in both of these cases the majority of the identified students had limited English proficiency). Given that many students identified in these groups participate in the assessment, the percentage of fourth-grade students who were excluded ranged between 7 and 10 in public schools across the nation, in central cities, and in four of the six public school districts included in this report. In contrast to this, in Atlanta only 2 percent of the fourth-grade students were excluded from the assessment because they had disabilities (1 percent) or were limited English proficient (1 percent). At the same time, 15 percent of the fourth-grade students in Houston were excluded from the assessment, with 13 percent excluded because they had limited English proficiency.

In the eighth grade, the percentage of students identified with disabilities and/or limited English proficiency ranged from 8 to 22 percent in public schools across the nation, in central cities, and two of the five districts. Atlanta was below the national average, with only 8 percent of its eighth-grade students identified in these groups; and Houston and Los Angeles were above the national average with 27 and 35 percent, respectively, of their eighth-grade students identified in these groups (in both of these cases the majority of the identified students had limited English proficiency). The percentages of eighth-grade students excluded from the assessment were more similar across jurisdictions. The exclusion rates for the nation, central city public schools, and four of the five districts ranged from 5 to 7 percent. In contrast, Atlanta excluded only 3 percent of its eighth-graders.

In both grades the percentage of students assessed with accommodations ranged from 0 or 1 percent to 11 percent across jurisdictions. The variability in the identification, exclusion, and accommodation rtes should be taken into consideration in interpreting the results and making comparisons (see appendix A, table A.1).

BEST COPY AVAILABLE

Evaluating Students' Writing

Student responses in the NAEP 2002 writing assessment were evaluated according to scoring guide criteria describing six performance levels. Scoring guides were developed for narrative, informative, and persuasive writing at each grade level. For example, the same scoring guide was used for all grade 8 narrative tasks. The guides included specific notes for raters that described various student approaches to the writing task and offered anchor or prototypical student responses at each grade level. For each task, a wide spectrum of student approaches was judged to be acceptable. Acknowledging developmental differences between fourth-, eighth-, and twelfth-graders, the scoring guides reflect higher performance expectations for higher-level students (for more details, see chapter 4 of the NAEP 2002 writing report).[7] Following the framework, the scoring guides emphasize students' abilities to develop and elaborate ideas, organize their thoughts, and write grammatically correct prose. The criteria for measuring command of written English mechanics differed by grade, but were the same across the three purposes for writing (narrative, informative, and persuasive) within each grade. Since a time-controlled writing context constrains students' opportunities to plan and revise, responses to assessment tasks were viewed as first drafts and evaluated accordingly.

Student responses to all 25-minute tasks were analyzed based on descriptive rubrics in the 6-level scoring guides for narrative, informative, and persuasive writing. A scale of 1 through 6 —Unsatisfactory, Insufficient, Uneven, Sufficient, Skillful, and Excellent—was used in evaluating individual student responses to writing tasks.

Reporting the Writing Assessment Results

Results from the NAEP writing assessment are presented in two ways: as scale scores on a scale from 0 to 300 and as percentages of students attaining achievement levels (*Basic, Proficient,* and *Advanced).* The scale scores, indicating how much students *know and can do* in writing, are presented as average scale scores and as scale scores at selected percentiles. In this report, the term *average scores* refers to average scores on the NAEP writing scale. The achievement level results provide further information by indicating the degree to which student performance meets the standards set for what they *should know and be able to do.* Results are reported only for groups or subgroups of students; an individual student's performance cannot be reported based on NAEP assessment.

Student responses to all 25-minute writing tasks were analyzed to determine the percentage of students scoring at each level on the 6-level guides for narrative, informative, and persuasive writing. The analysis entails summarizing the results on

[7] Persky, H. R., Daane, M. C., and Jin, Y. (2003). *The Nation's Report Card: Writing 2002.* NCES 2003-529. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

separate subscales for each writing purpose and then combining the separate scales to form a single composite writing scale. This analysis yields the overall scale of 0–300 for each grade (grades 4 and 8). Performance for each grade is scaled separately; therefore, average scale scores cannot be compared across grades. Thus, equal scores on grade 4 and grade 8 scales do not imply equal levels of writing ability. (For more information on scaling procedures, see "Data Analysis and Item Response Theory (IRT) Scaling" in appendix C of this report.)

Student writing performance is also reported in terms of three achievement levels: *Basic, Proficient,* and *Advanced.* The three achievement levels are defined as follows:

- *Basic:* This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- *Proficient:* This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- *Advanced:* This level signifies superior performance.

For reporting purposes, achievement level cut scores are placed on the writing scale to show the cutpoints between the *Basic, Proficient,* and *Advanced* levels. The achievement level results are then reported as percentages of students within each achievement level range, as well as the percentage of students at or above *Basic* and at or above *Proficient.*

The achievement levels are performance standards adopted by NAGB as part of its statutory responsibilities. The levels represent collective judgments of what students should know and be able to do for each grade assessed. They are based on recommendations from broadly representative panels of classroom teachers, education specialists, and members of the general public. As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that the achievement levels are to be used on a trial basis until it is determined that the achievement levels are "reasonable, valid, and informative to the public."[8] However, both NCES and NAGB believe these performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials as a common yardstick of academic performance. For a full discussion of achievement levels, see "The Setting of Achievement Levels" and "Trial Status of Achievement Levels" in chapter 1 of the NAEP 2002 writing report.[9]

---

[8] No Child Left behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

[9] Persky, H. R., Daane, M. C., and Jin, Y. (2003). *The Nation's Report Card: Writing 2002.* NCES 2003-529. Washington, DC: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Specific definitions of the *Basic*, *Proficient*, and *Advanced* writing achievement levels for grades 4 and 8 are presented in figures 1.2 and 1.3. The achievement levels are cumulative; therefore, students performing at the *Proficient* level also display the competencies associated with the *Basic* level, and students at the *Advanced* level also demonstrate the competencies associated with both the *Basic* and the *Proficient* levels. The score ranges for the NAGB achievement levels on the NAEP scale are as follows: Grade 4, *Basic:* 115–175; *Proficient:* 176–224; *Advanced:* 225 and above; Grade 8, *Basic:* 114–172; *Proficient:* 173–223; *Advanced:* 224 and above.

Interpreting NAEP Results

The average scores and percentages presented in this report are estimates based on samples of students rather than on entire populations. Moreover, the collection of writing tasks used at each grade level is but a sample of the many tasks that could have been used to assess the skills and abilities described in the NAEP writing framework. The results are subject to a measure of uncertainty, reflected in the standard error of the estimates—a range of a few points plus or minus the average score or percentage—which accounts for potential fluctuations in average scores or percentages that are due to sampling and measurement error. The estimated standard errors for the estimated scale scores and percentages in this report are accessible through the NAEP Data Tool on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/). Examples of these estimated standard errors are also provided in appendix C of this report.

The differences between scale scores and between percentages discussed in the following chapters take into account the standard errors associated with the estimates. Comparisons are based on statistical tests that consider both the magnitude of the difference between the group average scores or percentages and the standard errors of those statistics. Estimates based on smaller subgroups are likely to have relatively large standard errors. As a consequence, some seemingly large differences may not be statistically significant. That is, it cannot be determined whether these differences are due to the particular makeup of the samples of students who are selected or the true differences in the population of interest. When this is the case, the term "apparent difference" is used in this report. Differences between scores or between percentages are reported as such in this report only when they are significant from a statistical perspective. All differences reported are significant at the .05 level (with appropriate adjustments for multiple comparisons). The term "significant" is intended to identify statistically dependable differences in average scores or percentages and not to imply a judgment about the absolute magnitude or the educational relevance of the differences. Throughout

BEST COPY AVAILABLE

this report, comparisons are made among districts and between districts and public schools in the nation, as well as between districts and central city public schools in the nation. (See "NAEP Reporting Groups" in appendix C for details on how "central city public schools" was defined.) Statistical tests are used to determine whether the differences between average scores are significant. Only statistically significant differences are cited in this report.

The reader is cautioned to rely on the reported differences in the text and tables, which are statistically significant, rather than on the apparent magnitude of any differences. The standard errors are available on the NAEP web site (http://www.nces.ed.gov/nationsreportcard/naepdata).

Cautions in Interpretations

As described earlier, the NAEP writing scale makes it possible to examine relationships between students' performance and various background factors measured by NAEP; however, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not reflect the influence of unmeasured variables. The results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

Figure 1.2 Descriptions of NAEP writing achievement levels, grade 4

Grade 4 Achievement Levels

The following statements describe the kinds of things fourth-grade students should be able to do in writing at each level of achievement. These statements should be interpreted with the constraints of the National Assessment of Educational Progress (NAEP) in mind. Student performances reported with respect to these descriptions are in response to two age-appropriate writing tasks completed within 25 minutes each. Students are not advised of the writing tasks in advance nor engaged in pre-writing instruction and preparation; however, they are given a set of "ideas for planning and reviewing" their writing for the assessment. Although the Writing NAEP cannot fully assess students' abilities to produce a polished piece of writing, the results do provide valuable information about students' abilities to generate writing in response to a variety of purposes, tasks, and audiences within a rather limited period of time.

Basic (115)

Fourth-grade students performing at the Basic level should be able to produce a somewhat organized and detailed response within the time allowed that shows a general grasp of the writing task they have been assigned.

Fourth-grade students performing at the *Basic* level should be able to produce a somewhat organized response within the time allowed that shows a general grasp of the writing task they have been assigned. Their writing should include some supporting details. Its grammar, spelling, and capitalization should be accurate enough to communicate to a reader, although there may be mistakes that get in the way of meaning.

Proficient (176)

Fourth-grade students performing at the Proficient level should be able to produce an organized response within the time allowed that shows an understanding of the writing task they have been assigned. Their writing should include details that support and develop their main idea, and it should show that these students are aware of the audience they are expected to address.

Fourth-grade students performing at the *Proficient* level should be able to produce an organized response within the time allowed that shows an understanding of the writing task they have been assigned. Their writing should include details that support and develop the main idea of the piece, and its form, content, and language should show that these students are aware of the audience they are expected to address. The grammar, spelling, and capitalization in the work should be accurate enough to communicate to a reader; there may be some mistakes, but these should not get in the way of meaning.

Advanced (225)

Fourth-grade students performing at the Advanced level should be able to produce an effective, well developed response within the time allowed that shows a clear understanding of the writing task they have been assigned and the audience they are expected to address. Their writing should include details and be clearly organized, should use precise and varied language, and may show signs of analytical, evaluative, or creative thinking.

Fourth-grade students performing at the *Advanced* level should be able to produce an effective, well developed response within the time allowed that shows a clear understanding of the writing task they have been assigned. Their writing should be clearly organized, making use of techniques such as consistency in topic or theme, sequencing, and a clearly marked beginning and ending. It should make use of precise and varied language to speak to the audience the students are expected to address, and it should include details and elaboration that support and develop the main idea of the piece. Their writing may also show signs of analytical, evaluative, or creative thinking. The grammar, spelling, and capitalization in the work should be accurate enough to communicate clearly; mistakes should be so few and so minor that a reader can easily skim over them.

SOURCE: S. C. Loomis and M. L. Bourque (Eds.) (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Writing.* U.S. Dept. of Education, Office of Educational Research, National Center for Education Statistics. Washington, DC.

BEST COPY AVAILABLE

Figure 1.3 Descriptions of NAEP writing achievement levels, grade 8

Grade 8 Achievement Levels

The following statements describe the kinds of things eighth-grade students should be able to do in writing at each level of achievement. These statements should be interpreted with the constraints of the National Assessment of Educational Progress (NAEP) in mind. Student performances reported with respect to these descriptions are in response to two age-appropriate writing tasks completed within 25 minutes each. Students are not advised of the writing tasks in advance nor engaged in pre-writing instruction and preparation; however, they are given a set of "ideas for planning and reviewing" their writing for the assessment. Although the Writing NAEP cannot fully assess students' abilities to produce a polished piece of writing, the results do provide valuable information about students' abilities to generate writing in response to a variety of purposes, tasks, and audiences within a rather limited period of time.

Basic (114)

Eighth-grade students performing at the Basic level should be able to produce an effective response within the time allowed that shows a general understanding of the writing task they have been assigned. Their writing should show that these students are aware of the audience they are expected to address, and it should include supporting details in an organized way.

Eighth-grade students performing at the *Basic* level should be able to produce an effective response within the time allowed that shows a general understanding of the writing task they have been assigned. Their writing should show that these students are aware of the audience they are expected to address, and it should include supporting details in an organized way. The grammar, spelling, punctuation, and capitalization in the work should be accurate enough to communicate to a reader, although there may be mistakes that get in the way of meaning.

Proficient (173)

Eighth-grade students performing at the Proficient level should be able to produce a detailed and organized response within the time allowed that shows an understanding of both the writing task they have been assigned and the audience they are expected to address. Their writing should include precise language and varied sentence structure, and it may show analytical, evaluative, or creative thinking.

Eighth-grade students performing at the *Proficient* level should be able to produce an effective response within the time allowed that shows an understanding of both the writing task they have been assigned and the audience they are expected to address. Their writing should be organized, making use of techniques such as sequencing or a clearly marked beginning and ending, and it should make use of details and some elaboration to support and develop the main idea of the piece. Their writing should include precise language and some variety in sentence structure, and it may show analytical, evaluative, or creative thinking. The grammar, spelling, punctuation, and capitalization in the work should be accurate enough to communicate to a reader; there may be some errors, but these should not get in the way of meaning.

Advanced (224)

Eighth-grade students performing at the Advanced level should be able to produce a fully developed response within the time allowed that shows a clear understanding of both the writing task they have been assigned and the audience they are expected to address. Their writing should show some analytical, evaluative, or creative thinking and may make use of literary strategies to clarify a point. At the same time, the writing should be clearly organized, demonstrating precise word choice and varied sentence structure.

Eighth-grade students performing at the *Advanced* level should be able to produce an effective and fully developed response within the time allowed that shows a clear understanding of both the writing task they have been assigned and the audience they are expected to address. Their writing should show some analytical, evaluative, or creative thinking, and should demonstrate precise word choice and varied sentence structure. Their work should include details and elaboration that support and develop the main idea of the piece, and it may make use of strategies such as analogies, illustrations, examples, anecdotes, or figurative language to clarify a point. At the same time, the writing should show that these students can keep their work clearly and consistently organized. Writing by eighth-grade students performing at the *Advanced* level should contain few errors in grammar, spelling, punctuation, capitalization, and sentence structure. These writers should demonstrate good control of these elements and may use them for stylistic effect in their work.

SOURCE: S. C. Loomis and M. L. Bourque (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Writing.* U.S. Dept. of Education, Office of Educational Research, National Center for Education Statistics. Washington, DC.

BEST COPY AVAILABLE

2 Average Scale Score and Achievement Level Results for the Trial Urban District Assessment

This chapter presents the NAEP 2002 writing results for the assessed urban districts and the District of Columbia at grades 4 and 8. Performance is reported as average scores on the NAEP writing scale, which ranges from 0 to 300, and in terms of the three writing achievement levels—*Basic*, *Proficient* and *Advanced*.

Urban District Scale Score and Percentile Results

Table 2.1 shows the overall performance of fourth- and eighth-grade students in the urban districts that participated in the 2002 writing assessment. In order to provide a context for these data, table 2.1 also displays the results for students attending public schools in the nation as a whole, as well as for public schools located in central cities across the nation. In table 2.1 and subsequent tables and figures in this report, a double asterisk (**) marks district statistics (average scores or percentages) that were found to be significantly different from the comparable statistic in public schools in the nation, and a single asterisk (*) marks district statistics that were significantly different from those of public schools in central cities. Following standards established by the Federal Office of Management and Budget, the U.S. Census Bureau defines a central city as a city of 50,000 people or more that is the largest in its metropolitan area, or can otherwise be regarded as "central," taking into account such factors as commuting patterns. (See appendix C for more detailed definitions of geographical areas). The districts that participated in

the Trial Urban Assessment, Atlanta City, Chicago School District 299, Houston ISD, Los Angeles Unified, and New York City Public Schools, and Washington, DC are all located in central cities. The term means "a city that is central", not "the central part of a city" or the "inner city." Central cities are defined for this report as including large and mid-size cities. All students in these districts, except for Houston and Los Angeles attended schools in central cities. In the Houston and Los Angeles districts, some students included in this study attended schools located in the urban fringe. These included 5 percent and 19 percent of fourth-grade students in Houston and Los Angeles respectively, as well as 24 percent of eighth-grade students in Los Angeles.

The first column in the table presents the average score by district on the NAEP writing scale. Results at each grade are scaled independently; therefore, cross-grade score comparisons cannot be made. The average score for fourth-grade students in public schools in the nation was 153 and for students in central city schools was 147. The average score for eighth-grade students in public schools in the nation was 152 and in central city schools was 143. At both grades, the average scores for students in the nation and for central city public schools generally were higher than those in these urban districts; exceptions to this generalization were found for New York City and Houston. Fourth-grade students in New York City had higher average scores than fourth-graders in central city schools, while no significant differences were found between average scores for fourth- and eighth-grade students in Houston and those in central city schools.

Details on the statistical significance of these performance differences are provided below in the "Comparisons Among Districts by Average Scale Scores" section. Data for New York City at grade 8 do not appear in this report because the district did not meet the required 70 percent participation rate (see appendix C, "Standards for Sample Participation and Reporting of Results").

The remaining columns in table 2.1 indicate the scores attained on the 0-300 writing scale of students at selected percentiles. Each percentile indicates the percentage of students whose scores fell below a particular point on the NAEP writing scale. For example, in Atlanta the 75th percentile score at grade 4 was 161 in 2002, indicating that 75 percent of fourth-graders scored below 161. Table 2.1 shows the writing scale score for students scoring at the 25th, 50th, and 75th percentiles. An examination of scores at different percentiles at each grade indicates the score differences between higher-, middle-, and lower-performing students within a district and thus provides more detailed information than does the average score alone. The corresponding standard errors for these percentile scores are displayed in table C.6 in appendix C.

When NAEP conducts a second Trial Urban District Assessment in writing, comparing the score values for percentiles across years will be useful in pinpointing where changes may have occurred within a district's score distribution. For this report, the comparison of the scores at various percentiles across districts gives a rough indication of how score distributions may differ.

Table 2.1 Average writing scale scores and selected percentiles, grades 4 and 8 public schools: By urban district, 2002

	Scale score distribution			
	Average scale score	25th percentile	50th percentile	75th percentile
Grade 4				
Nation (Public)	153	128	153	178
Central city (Public) [1]	147 **	122 **	146 **	171 **
Atlanta	140 *,**	117 *,**	139 *,**	161 *,**
Chicago	138 *,**	116 *,**	137 *,**	160 *,**
District of Columbia	135 *,**	113 *,**	134 *,**	157 *,**
Houston	148	123	147	174
Los Angeles	141 *,**	117 **	141 *,**	165 *,**
New York City [1]	153 *	128	154 *	178
Grade 8				
Nation (Public)	152	127	153	178
Central city (Public) [2]	143 **	117 **	144 **	170 **
Atlanta	130 *,**	107 *,**	129 *,**	151 *,**
Chicago	136 *,**	111 **	136 *,**	160 *,**
District of Columbia	128 *,**	105 *,**	128 *,**	152 *,**
Houston	138 **	113 **	139 **	165 **
Los Angeles	128 *,**	104 *,**	128 *,**	152 *,**

[1] Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.

* Significantly different from central city public schools.

** Significantly different from nation (public schools).

[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.

[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Comparisons Among Districts by Average Scale Scores

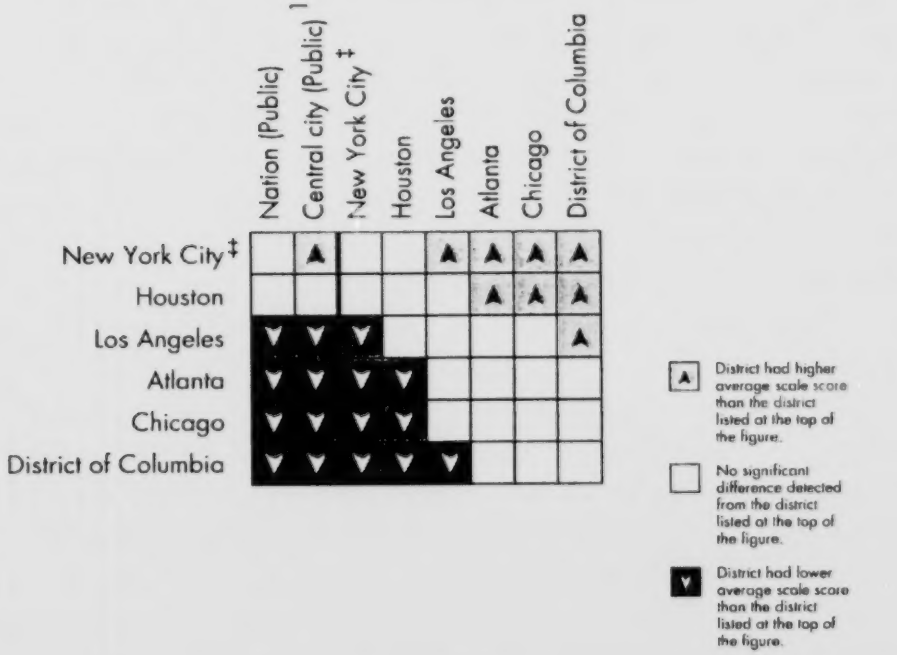
Figures 2.1 and 2.2 display the results of statistical significance tests of differences in NAEP 2002 average writing scores between any two urban districts at grade 4 and at grade 8, respectively. These figures are similar to mileage charts on travel maps. Read across the row corresponding to a district to the column of the jurisdiction chosen for comparison. If the resulting cell is not shaded, the difference between the two average scores was not found to be statistically significant. If the cell is lightly shaded (with an upward arrow), the average scale score of the district on the row on the left was higher than that of the jurisdiction named at the top of the column. The darkly shaded cells (with a downward arrow) indicate that the average scale score of the district on the rows on the left was lower than that of the jurisdiction selected at the top of the column. For example, in figure 2.1, the leftmost cell in the fourth row compares the average score at grade 4 in Atlanta to the average score in the nation. The shading and the arrow in this cell indicate that the average score in Atlanta was lower than the national average score and that the difference was statistically significant.

At grade 4, the average score for students in the nation was higher than the average score in each of the districts except New York City and Houston, where the average scores were not found to be significantly different from the national average score. Fourth-grade average scores were lower in most districts than the average score in the central city schools except in New York City, where the average score was higher, and in Houston, where the average score was not found to differ significantly. At grade 4, the average score for students in New York City was higher than those in all the other districts except Houston. The average score for students in Houston was higher than average scores in Atlanta, Chicago, and the District of Columbia, but not found to differ significantly from average scores in Los Angeles and New York City. The average score for students in Los Angeles was higher than the average score in the District of Columbia, but not found to differ significantly from the average scores in Atlanta and Chicago. Other apparent differences were not determined to be statistically significant.

Figure 2.1 Cross-district comparisons of average writing scale scores, grade 4 public schools: By urban district, 2002

Grade 4

Instructions: Read across the row corresponding to an urban district listed to the left of the chart. Match the shading intensity to the key below to determine whether the average writing scale score of this district was found to be higher than, not significantly different from, or lower than that of the entity named in the column heading. For example, in the row for Atlanta, Atlanta's average score was lower than the average scores in the nation, central cities, New York City and Houston, but not found to differ significantly from the average scores in Los Angeles, Chicago, and the District of Columbia.


Nation (Public)
Central city (Public) 1
New York City ‡
Houston
Los Angeles
Atlanta
Chicago
District of Columbia
New York City ‡
Houston
Los Angeles
Atlanta
Chicago
District of Columbia
District had higher average scale score than the district listed at the top of the figure.
No significant difference detected from the district listed at the top of the figure.
District had lower average scale score than the district listed at the top of the figure.

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.

1 For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

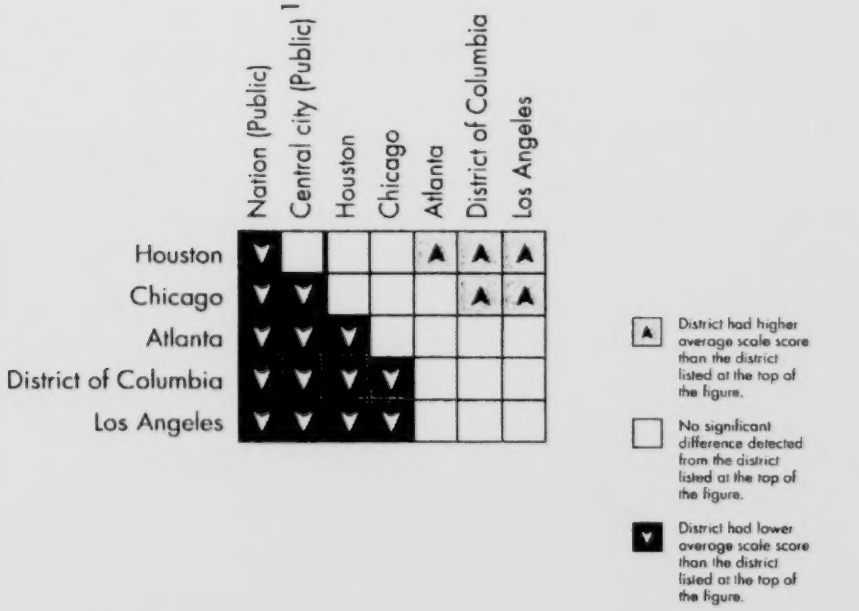
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Figure 2.2 Cross-district comparisons of average writing scale scores, grade 8 public schools: By urban district, 2002

Grade 8

Instructions: Read across the row corresponding to an urban district listed to the left of the chart. Match the shading intensity to the key below to determine whether the average writing scale score of this district was found to be higher than, not significantly different from, or lower than the entity named in the column heading. For example, in the row for the District of Columbia, the average score for the District of Columbia was lower than the average scores for the nation, central city, Houston and Chicago, and not found to differ significantly from Atlanta and Los Angeles.


Nation (Public)
Central city (Public) [1]
Houston
Chicago
Atlanta
District of Columbia
Los Angeles
Houston
Chicago
Atlanta
District of Columbia
Los Angeles
District had higher average scale score than the district listed at the top of the figure.
No significant difference detected from the district listed at the top of the figure.
District had lower average scale score than the district listed at the top of the figure.

[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

As shown in figure 2.2, eighth-grade students' average scores in all five districts were lower than the national average and lower than the central city average, with the exception of Houston, where the average score was not found to be significantly different from the average score for central city schools. At grade 8, the average score in Houston was higher than the average scores in Atlanta, the District of Columbia, and Los Angeles. The average score in Chicago was higher than the average scores in the District of Columbia and Los Angeles and not found to differ significantly from the scores in Houston and Atlanta. Other apparent differences were not determined to be statistically significant.

BEST COPY AVAILABLE

Writing Achievement Level Results

In addition to reporting average writing scale scores, NAEP reports writing performance by achievement levels. The writing achievement levels are *Basic*, *Proficient*, and *Advanced*. The setting of the achievement levels is discussed in chapter 1.

Table 2.2 presents the percentages of students at grades 4 and 8 who performed below the *Basic* level, at or above the *Basic* level, at or above the *Proficient* level, and at the *Advanced* level.

At grade 4, the percentage of students performing at or above *Proficient* ranged from 11 percent in the District of Columbia to 27 percent in New York City.

Table 2.2 Percentage of students at or above each writing achievement level, grades 4 and 8 public schools: By urban district, 2002

	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Grade 4				
Nation (Public)	15	85	27	2
Central city (Public) [1]	19 **	81 **	21 **	2 **
Atlanta	23 **	77 **	13 *,**	1
Chicago	24 *,**	76 *,**	12 *,**	#
District of Columbia	27 *,**	73 *,**	11 *,**	1 *,**
Houston	19	81	23	2
Los Angeles	23 **	77 **	16 *,**	1 *,**
New York City [†]	15	85	27	2
Grade 8				
Nation (Public)	16	84	30	2
Central city (Public) [2]	23 **	77 **	22 **	1 **
Atlanta	32 *,**	68 *,**	10 *,**	#
Chicago	28 *,**	72 *,**	16 *,**	1
District of Columbia	34 *,**	66 *,**	10 *,**	#
Houston	26 **	74 **	19 **	1
Los Angeles	36 *,**	64 *,**	11 *,**	#

Percentage rounds to zero.
† Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.
[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Twenty-seven percent of the students across the nation performed at or above the *Proficient* level, while 21 percent of all central city students reached this level of performance. The percentages of fourth-grade students reaching at least the *Basic* achievement level ranged from 73 percent in the District of Columbia to 85 percent in New York City. In the nation as a whole, 85 percent of students and 81 percent of central-city students performed at or above the *Basic* level.

At grade 8, the percentages of students at or above *Proficient* ranged from 10 percent in Atlanta and the District of Columbia to 19 percent in Houston, while 30 percent of students in public schools in the nation and 22 percent in central city public schools performed at or above the *Proficient* level. At grade 8, the percentages of students performing at or above *Basic* ranged from 64 percent in Los Angeles to 74 percent in Houston. In the national public schools, 84 percent of students performed at or above *Basic*, and 77 percent of central city students performed at the same level.

Comparisons Among Districts by Achievement Levels

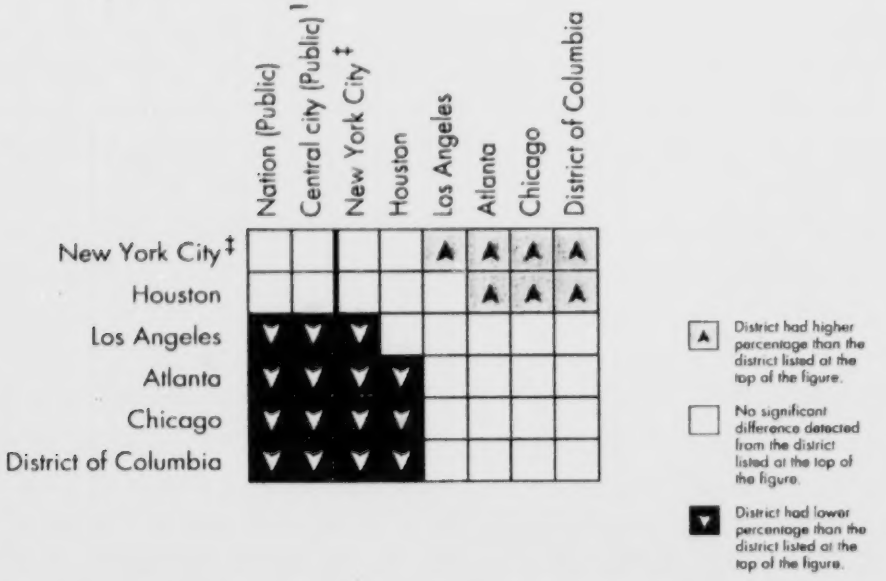
Figures 2.3 and 2.4 are "mileage charts" similar to figures 2.1 and 2.2; in this case the data compared represent percentages of students performing at or above the *Proficient* level. At grade 4, the percentages of students performing at or above *Proficient* in public schools nationally and in public central city schools were higher than the percentages in all districts except Houston and New York City, where the differences were not found to be statistically significant. At grade 4, the percentage of students performing at or above *Proficient* in New York City was higher than the percentages in four other districts and not found to differ significantly from the percentage in Houston.

While not displayed in the figures, there were also interesting patterns in the percentages of students performing at or above the *Basic* level. No significant difference was found between the percentages of fourth-grade students performing at or above *Basic* in Houston and New York City and the percentages of fourth-graders performing at the same level in the nation and central cities. The percentages of fourth-grade students performing at or above *Basic* were lower in Chicago and the District of Columbia than the percentages of students performing at or above *Basic* in central cities. Fourth-grade students in Houston and New York City were more likely than those in the District of Columbia to perform at or above the *Basic* level. Other apparent differences were not found to be statistically significant.

Figure 2.3 Cross-district comparisons of percentage of students at or above *Proficient* in writing, grade 4 public schools: By urban district, 2002

Grade 4

Instructions: Read across the row corresponding to an urban district listed to the left of the chart. Match the shading intensity to the key below to determine whether the percentage of students at or above *Proficient* in this district was found to be higher than, not significantly different from, or lower than the district or jurisdiction in the column heading. For example, in the row for Chicago: The percentage of students at or above *Proficient* in Chicago was lower than the percentage in the nation, central cities, New York City, and Houston, but not found to differ significantly from the percentage in Los Angeles, Atlanta, and the District of Columbia.


Nation (Public)
Central city (Public)[1]
New York City‡
Houston
Los Angeles
Atlanta
Chicago
District of Columbia
New York City‡
Houston
Los Angeles
Atlanta
Chicago
District of Columbia
District had higher percentage than the district listed at the top of the figure.
No significant difference detected from the district listed at the top of the figure.
District had lower percentage than the district listed at the top of the figure.

‡ Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.

[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

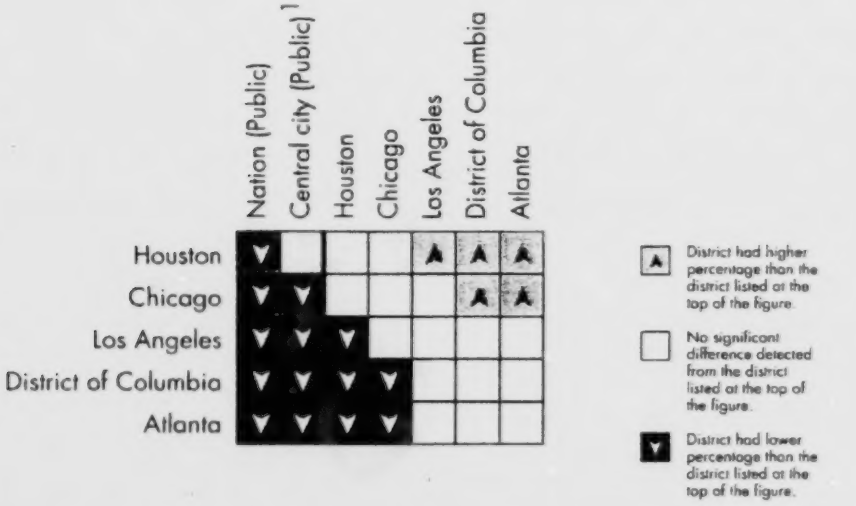
At grade 8, the percentage of students performing at or above *Proficient* was higher for the nation than for all urban districts assessed, and higher for central cities than for all urban districts assessed except Houston. No significant difference was detected between the percentages of students performing at or above *Proficient* in Houston and central cities. The percentages of eighth-graders performing at or above *Proficient* in Chicago and Houston were not found to differ significantly from each other, and both were higher than the comparable percentages in the District of Columbia and Atlanta.

The percentage of students performing at or above *Basic* was higher in public schools in the nation than in all five districts reported and higher in the central city schools than in all these districts except Houston, which did not differ significantly from the central cities on this performance measure. Students in Houston were more likely than those in the District of Columbia and Los Angeles to perform at or above the *Basic* level.

Figure 2.4 Cross-district comparisons of percentage of students at or above *Proficient* in writing, grade 8 public schools: By urban district, 2002

Grade 8

Instructions: Read across the row corresponding to an urban district listed to the left of the chart. Match the shading intensity to the key below to determine whether the percentage of students at or above *Proficient* in this district was found to be higher than, not significantly different from, or lower than the district or jurisdiction in the column heading. For example, in the row for Houston: The percentage of students at or above *Proficient* in Houston was lower than in the nation, was not found to differ from the central cities and Chicago, and was higher than Los Angeles, the District of Columbia, and Atlanta.


Nation (Public)
Central city (Public)[1]
Houston
Chicago
Los Angeles
District of Columbia
Atlanta
Houston
Chicago
Los Angeles
District of Columbia
Atlanta
District had higher percentage than the district listed at the top of the figure.
No significant difference detected from the district listed at the top of the figure.
District had lower percentage than the district listed at the top of the figure.

[1] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.

NOTE: The between-district comparisons take into account sampling and measurement error and that each district is being compared with every other district shown. Significance is determined by an application of a multiple-comparison procedure.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

3 Results for Subgroups

In addition to reporting on the performance of all students, NAEP provides results for a variety of subgroups of students for each grade level assessed. The subgroup results show how these groups of students performed in comparison with one another. When additional years of urban district assessments are conducted, these results will provide a basis for examining each group's progress over time.

This chapter includes average writing scores and achievement level results at grade 4 for subgroups of students in six urban districts—the five that participated in the Trial Urban District Assessment and the District of Columbia—and results for grade 8 in Atlanta, Chicago, Houston, Los Angeles, and the District of Columbia. New York City data for grade 8 are not reported because participation did not meet reporting criteria. Results are reported by gender, race/ethnicity, students' eligibility for free/reduced-price lunch, and by parents' highest level of education.

Throughout this chapter, student subgroup results in each district are also compared to the results for central city schools. In most cases, the average scores and achievement level results for central city schools were below those for the nation. All differences noted as such in this chapter are statistically significant. In interpreting the results, it is important to bear in mind that the estimated average score for a particular group does not include the whole range of performance within that group. Differences in subgroup performance cannot be ascribed solely to students' membership in an identified subgroup. Average student performance is affected by the interaction of a complex set of factors not discussed in this report or addressed by NAEP assessments.

Performance of Selected Subgroups in the Trial Urban District Assessment

Gender

Educators and government agencies have produced a body of research rich in data documenting gender differences in language arts achievement.[1] National results for the NAEP 1998 and 2002 writing assessments indicate that female students outperformed male students at grades 4, 8, and 12.

Table 3.1 shows average writing scores and achievement level results by gender at grade 4. Female fourth-graders had a higher average score than male fourth-graders in each of the urban districts. Average scores for male students at grade 4 ranged from 127 in the District of Columbia to 145 in New York City. No significant difference was detected between the average scores for male fourth-grade students in Houston and New York City and their counterparts in central city schools. The average scores of male fourth-graders in Atlanta, Chicago, the District of Columbia, and Los Angeles were lower than the average score for male fourth-grade students in central city schools.

The average scores for female students at grade 4 in the six districts ranged from 143 in the District of Columbia to 160 in New York City. No significant difference was detected between the average scores for female fourth-grade students in Houston and New York City and the average score for female fourth-graders in the central city schools. Female fourth-graders in Atlanta, Chicago, the District of Columbia, and Los Angeles had lower average scores than female fourth-grade students in central city schools.

---

[1] U.S. Department of Education. (2001). *Reading for Understanding: Towards an R & D Program in Reading Comprehension*. Washington, DC: Author.

MacMillan, P. (2000). Simultaneous Measurement of Reading Growth, Gender, and Relative-Age Effects: Many Faceted Rasch Applied to CBM Reading Scores. *Journal of Applied Measurement, 1*(4), 393–408.

Peterson, S. (2001). Gender Identities and Self-Expression in Classroom Narrative Writing. *Language Arts, 18*(5), 451–439.

Thomas, P. (1994). Writing, Reading, and Gender. *Gifted Education International, 9*(3), 154–158.

BEST COPY AVAILABLE

Table 3.1 Average writing scale scores and percentage of students at or above each achievement level, by gender, grade 4 public schools: By urban district, 2002

Grade 4		Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male							
	Nation (Public)	51	144	20	80	18	1
	Central city (Public) [1]	51	139 **	24 **	76 **	15 **	1
	Atlanta	47	131 *,**	31 *,**	69 *,**	8 *,**	1
	Chicago	50	130 *,**	32 *,**	68 *,**	7 *,**	#
	District of Columbia	49	127 *,**	37 *,**	63 *,**	7 *,**	# *,**
	Houston	48	141	24	76	17	1
	Los Angeles	51	134 *,**	29 **	71 **	11 *,**	#
	New York City !	50	145	21	79	21	1
Female							
	Nation (Public)	49	162	10	90	35	3
	Central city (Public) [1]	49	155 **	13 **	87 **	27 **	2 **
	Atlanta	53	147 *,**	17 **	83 **	18 *,**	2
	Chicago	50	145 *,**	17 **	83 **	16 *,**	#
	District of Columbia	51	143 *,**	18 *,**	82 *,**	15 *,**	1 *,**
	Houston	52	154 **	15 **	85 **	28 **	3
	Los Angeles	49	148 *,**	16 **	84 **	21 *,**	1 *,**
	New York City !	50	160	10	90	33	3

Percentage rounds to zero.
! Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Table 3.2 Average writing scale scores and percentage of students at or above each achievement level, by gender, grade 8 public schools: By urban district, 2002

Grade 8	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male						
Nation (Public)	50	141	23	77	20	1
Central city (Public) [1]	50	133 **	30 **	70 **	14 **	1
Atlanta	49	122 *,**	40 *,**	60 *,**	6 *,**	#
Chicago	48	126 *,**	37 **	63 **	11 **	1
District of Columbia	49	120 *,**	43 *,**	57 *,**	6 *,**	#
Houston	47	124 *,**	37 **	63 **	9 *,**	#
Los Angeles	50	121 *,**	42 *,**	58 *,**	6 *,**	#
Female						
Nation (Public)	50	162	9	91	40	3
Central city (Public) [1]	50	153 **	15 **	85 **	30 **	2 **
Atlanta	51	137 *,**	24 *,**	76 *,**	13 *,**	1
Chicago	52	145 *,**	19 **	81 **	21 *,**	2
District of Columbia	51	136 *,**	25 *,**	75 *,**	14 *,**	#
Houston	53	150 **	16 **	84 **	28 **	1
Los Angeles	50	134 *,**	29 *,**	71 *,**	15 *,**	#

Percentage rounds to zero.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
[1] For comparison, at eighth grade 61 percent of students in central city public schools and 26 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

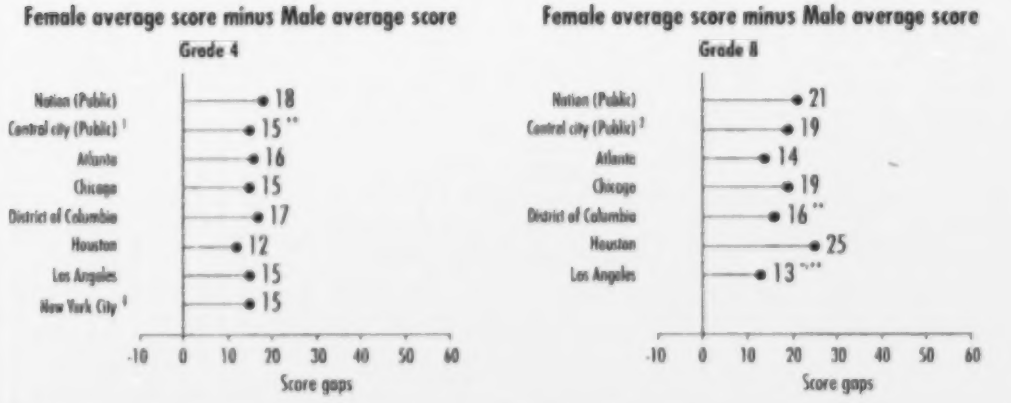
Table 3.2 displays average scores and achievement level results by gender for grade 8. In each of the five districts reported, the average score of female students was higher than the average score of male students. Average scores for male eighth-grade students ranged from 120 in the District of Columbia to 126 in Chicago. The average score for eighth-grade male students in central city schools was higher than that in each of the five urban districts reported.

For female students at grade 8 in the five districts, average scores ranged from 134 in Los Angeles to 150 in Houston. For female eighth-graders, the average score was higher in central city schools than in any of the districts except Houston: no significant difference was detected between the average score for female eighth-graders in Houston and in the central city schools.

BEST COPY AVAILABLE

An additional way to compare the performance of subgroups of students is to focus on the size of the difference between the average scores for the subgroups. Where the average scores of one subgroup have been consistently lower than those of another over time, the differences have come to be called "gaps." Figure 3.1 presents the "gaps" in average writing scores between female and male students in each district, in the central city schools, and in the nation. At grade 4, the gaps between the average scores of female and male students in the trial urban districts were not found to differ significantly from comparable gaps in public schools in the nation and in central cities. At grade 8, the score gap between female and male students was narrower in the District of Columbia than in public schools in the nation. In Los Angeles, the gender score gap was narrower than the comparable gaps in public schools in both central cites and in the nation.

Figure 3.1 Gaps in average writing scale scores, by gender, grades 4 and 8 public schools: By urban district, 2002



1 Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
1 For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.
2 For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Score gaps are calculated based on differences between unrounded average scale scores.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Race/Ethnicity

For the purpose of studying the progress of subgroups, NAEP collects information from school records on the racial/ethnic identification that best describes each participating student. The six mutually exclusive categories are White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. This information was the primary contributor to the classifications appearing below. For further details, see "NAEP Reporting Groups" in appendix C.

Tables 3.3 and 3.4 show average scores and achievement level results by racial and ethnic group membership for public-school students in the urban districts at grades 4 and 8, respectively. Only the race/ethnicity categories with sufficient membership to meet reporting requirements in the urban districts are reported below.

The distribution of students in terms of race/ethnicity in the urban districts differs from the distribution in the nation's public schools. Black or Hispanic students constitute the majority or the largest fraction of students in each of the urban districts in the trial assessment. Black students constitute over 85 percent in Atlanta and in the District of Columbia at fourth and eighth grades, as well as 47 percent of Chicago's fourth grade and half of its eighth grade. Hispanic students constitute a majority at fourth and eighth grades in Los Angeles and at eighth grade in Houston, as well as 47 percent of the fourth grade in Houston and 43 percent of the fourth grade in New York City. The highest percentage of White students in the urban districts is 15 percent in New York at grade 4 and 11 percent in Chicago at grade 8, whereas in the national writing assessment White students constitute majorities at both fourth and eighth grades: 60 percent at grade 4 and 64 percent at grade 8.

In five of the six urban districts in which a reliable comparison could be made, White fourth-graders had higher average scores than their Black and Hispanic counterparts. The White/Hispanic comparison in Atlanta could not be tested for statistical significance due to insufficient sample size. No significant difference was detected between the average scores of Asian American/Pacific Islander students and their White counterparts in Los Angeles and New York City.

BEST COPY AVAILABLE

Compared with Black students in fourth grade in the central cities, Black fourth-graders in Houston and New York City had higher average scores, while Black fourth-graders in the District of Columbia had a lower average score. Compared to the average score of Hispanic fourth-graders in the central city schools, no significant difference was detected for the average score of Hispanic fourth-graders in five of the six districts. The average score for Hispanic fourth-graders in Los Angeles was lower than that in central cities taken as a whole. Average scores for White students at grade 4 were higher in Atlanta, the District of Columbia, and New York City than for White students in the central city schools, and no significant differences were detected between the average scores of White students in Chicago, Houston, and Los Angeles and the average score in the central city schools. Finally, average scores for Asian/Pacific Islander students in Los Angeles and New York City were not found to be significantly different from the average score for their counterparts in central cities. Sample sizes in the other districts were insufficient for reliable significance testing.

Table A.3 Students with disabilities and limited English proficient students assessed with accommodations, by type of primary accommodation, grade 8: By urban district, 2002

	Weighted percentage of students sampled					
	Large-print book	Extended time	Small group	One-on-one	Scribe/ computer	Other
SD[1] and/or LEP[2] students						
Nation (Public)	0.03	1.89	2.86	0.11	0.02	0.06
Central city (Public) [3]	0.01	2.38	2.21	0.09	0.01	0.13
Atlanta	0.09	#	0.63	#	#	#
Chicago	#	1.70	5.39	#	#	#
District of Columbia	#	6.88	3.19	#	#	#
Houston	#	#	#	0.17	#	#
Los Angeles	#	1.38	1.63	0.06	#	0.06
SD[1] students						
Nation (Public)	0.03	1.68	2.76	0.11	0.02	0.06
Central city (Public) [3]	0.01	2.03	2.09	0.09	0.01	0.13
Atlanta	#	#	0.56	#	#	#
Chicago	#	1.55	5.19	#	#	#
District of Columbia	#	4.91	2.92	#	#	#
Houston	#	#	#	0.17	#	#
Los Angeles	#	1.38	1.56	0.06	#	0.06
LEP[2] students						
Nation (Public)	#	0.37	0.26	0.01	#	0.01
Central city (Public) [3]	#	0.71	0.29	0.02	#	0.04
Atlanta	0.09	#	0.07	#	#	#
Chicago	#	0.19	0.41	#	#	#
District of Columbia	#	2.19	0.38	#	#	#
Houston	#	#	#	#	#	#
Los Angeles	#	0.57	1.07	#	#	0.06

Percentage rounds to zero.
[1] Students with disabilities
[2] Limited English proficient students.
[3] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment.

BEST COPY AVAILABLE

The NAEP assessment may not allow some accommodations that are permitted in certain states. For example, in the writing assessment, translations of the prompts into languages other than English were not provided.

Investigating the Potential Effects of Exclusion Rates on Assessment Results

The variation in exclusion rates among states, jurisdictions, and school districts may threaten the comparison of results within a given year, because the results for different districts, states, or jurisdictions are based on different proportions of the populations.

NCES has funded research investigating ways in which excluded students might be included in the estimation of scores for total populations. NCES has also commissioned studies on the impact of assessment accommodations on overall scores. Several statistical approaches for estimating full populations (including estimates for excluded students) have been proposed using data from NAEP subjects, but none has yet been judged completely adequate. Regarding the impact of assessment accommodations on overall scores, NCES has conducted differential item functioning (DIF) studies of items assessed with and without accommodations.[3] In these studies, researchers found little evidence that accommodations changed the functioning of test questions.

BEST COPY AVAILABLE

[3] For information on DIF studies of items assessed with accommodations in the 1996 mathematics and science assessments, see Mazzeo, J. M., Carlson, J. E., Voelkl, K. E., and Lutkus, A. D. (1999). *Increasing the Participation of Special Needs Students in NAEP: A Report on 1996 NAEP Research Activities*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Appendix B
District-Level Contextual Variables

To help place results from the NAEP 2002 Trial Urban District Assessment into context, this appendix presents selected district-level data from sources other than NAEP.

These data are taken from the *Public Elementary/Secondary School Universe Survey, 2000–01*, and *Local Education Agency Universe Survey, 2000–01*. The interested reader can access most of the data presented in this appendix in the NCES report, *Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States* at http://nces.ed.gov/pubs2001/100_largest/tab_fig.asp.

Table B.1 Number of students, high-school completers, teachers, and schools, from non-NAEP sources: By urban district, school year 2000–01

	Number of students[1]	Number of 1999–2000 completers[2]	Number of full-time equivalent teachers	Number of schools
National	47,086,931	2,548,076	2,841,677	93,344
Central city	13,523,126	610,467	808,288	22,310
Atlanta	58,230	2,056	3,950	98
Chicago	435,261	14,875	23,935	602
District of Columbia	68,925	2,916	5,044	165
Houston	208,462	7,735	11,197	289
Los Angeles	721,346	27,439	35,150	659
New York City	1,066,516	40,827	65,242	1,213

[1] Count of students receiving educational services from school district may differ somewhat from the counts in table B.4, which reflect the count of students from the schools aggregated up to the school district.
[2] Includes high school diploma recipients as well as other high school completers (e.g., certificates of attendance), but does not include high school equivalencies (ex. GEDs).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.2 Poverty rate and federal funding, from non-NAEP sources: By urban district, 1996–97, 1997–98, 2001–02

	5- to 17-year olds, 1999–2000		Revenue, 1999–2000 (in thousands)				Title I allocations, 2001–02 (in thousands)	
	Population	Percent in poverty	Total	Federal	Federal as percent of total	Federal revenue per student	Basic grants	Concentration grants
Atlanta	66,131	34.6	$695,919	$56,558	8.1	$95.17	$752	$20,824
Chicago	540,667	24.8	3,604,873	539,567	15	124.97	963	137,865
District of Columbia	82,456	27.8	881,423	184,825	21	261.19	1,509	21,341
Houston	230,514	26	1,469,074	152,679	10.4	72.8	585	50,084
Los Angeles	863,656	28.3	5,757,819	592,508	10.3	83.45	668	178,088
New York City	1,397,739	28.3	10,945,650	1,119,944	10.2	104.11	812	397,541

NOTE: Federal revenue per student based on fall enrollment collected by the Bureau of the Census. Detail may not sum to totals, because of to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data survey; and U.S. Department of Commerce, "Survey of Local Government Finances."

BEST COPY AVAILABLE

Table B.3 Number of public elementary and secondary schools, by type of school, from non-NAEP sources: By urban district, school year 2000–01

	Total number of schools	Type of school [1]			
		Regular	Special education	Vocational education	Alternative
National	93,344	85,493	2,008	1,025	4,818
Central city	22,310	20,141	536	173	1,460
Atlanta	98	95	0	0	3
Chicago	602	578	24	0	0
District of Columbia	165	150	10	0	5
Houston	289	264	1	3	21
Los Angeles	659	582	18	0	59
New York City	1,213	1,106	8	18	81

[1] Type of school is a mutually exclusive category on the Common Core of Data. There are cases in which special education, vocational education, and alternative programs reside in other types of schools.
NOTE: Types of schools are defined in the following way on the Common Core of Data: Regular school — A public elementary/secondary school that does not focus primarily on vocational, special, or alternative education. NAEP is conducted only in regular schools. Special education school — A public elementary/secondary school that (a) focuses primarily on special education, including instruction for any of the following: hard of hearing, deaf, speech-impaired, health-impaired, orthopedically impaired, mentally retarded, seriously emotionally disturbed, multi-handicapped, visually handicapped, deaf and blind; and (b) adapts curriculum, materials, or instruction for students served. Vocational education school — A public elementary/secondary school that focuses primarily on vocational education, and provides education and training in one or more semi-skilled or technical operations. Alternative education school — A public elementary/secondary school that (a) addresses the needs of students that typically cannot be met in a regular school; (b) provides nontraditional education; (c) serves as an adjunct to a regular school; and (d) falls outside of the categories of regular, special education, or vocational education.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a.

BEST COPY AVAILABLE

Table B.4 Number of students in public elementary and secondary schools, by type of school, from non-NAEP sources: By urban district, school year 2000–01

		Type of school[1]			
	Total number of students[2]	Regular	Special education	Vocational education	Other and alternative
National	47,094,888	46,228,904	174,577	199,669	491,738
Central city	13,522,154	13,152,151	73,387	75,953	220,663
Atlanta	58,230	56,896	†	†	1,334
Chicago	435,261	431,553	3,708	†	†
District of Columbia	68,925	65,285	2,772	†	868
Houston	208,462	204,042	19	906	3,495
Los Angeles	721,346	704,932	4,480	†	11,934
New York City	1,066,945	1,009,319	1,688	25,409	30,529

† No students reported in membership for this type of school.

[1] Type of school is a mutually exclusive category on the Common Core of Data. There are cases in which special education, vocational education, and alternative programs reside in other types of schools.

[2] Student distribution by type of school is based on membership in the schools of the school district. Counts may vary from those in table B.1.

NOTE: Types of schools are defined in the following way on the Common Core of Data: Regular school — A public elementary/secondary school that does not focus primarily on vocational, special, or alternative education. NAEP is conducted only in regular schools. Special education school — A public elementary/secondary school that (a) focuses primarily on special education, including instruction for any of the following: hard of hearing, deaf, speech-impaired, health-impaired, orthopedically impaired, mentally retarded, seriously emotionally disturbed, multi-handicapped, visually handicapped, deaf and blind; and (b) adapts curriculum, materials, or instruction for students served. Vocational education school — A public elementary/secondary school that focuses primarily on vocational education, and provides education and training in one or more semi-skilled or technical operations. Alternative education school — A public elementary/secondary school that (a) addresses the needs of students that typically cannot be met in a regular school; (b) provides nontraditional education; (c) serves as an adjunct to a regular school; and (d) falls outside of the categories of regular, special education, or vocational education.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

BEST COPY AVAILABLE

Table B.5 Percentage of minority students, and number of public elementary and secondary schools within specified ranges of minority student percentages, from non-NAEP sources: By urban district, school year 2000–01

	Percentage of minority students enrolled	Number of schools with students [2]	Number of schools with a minority student [1] percentage of				
			0–20 percent	21–40 percent	41–60 percent	61–80 percent	81–100 percent
National	39.0	89,110	43,827	14,023	9,866	7,142	14,252
Central city	63.0	21,513	3,534	3,165	3,370	3,088	8,356
Atlanta	93.2	98	2	2	2	5	87
Chicago	90.4	596	1	20	26	51	498
District of Columbia	95.5	165	0	5	3	5	152
Houston	90.0	289	1	4	17	20	247
Los Angeles	90.1	659	4	16	50	88	501
New York City	84.7	1,207	29	39	114	139	886

[1] Minority students, as this table, includes all race/ethnicity categories except White, non-Hispanic.
[2] Includes only schools for which student membership by race/ethnicity was reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.6 Number of public elementary and secondary schools, by instructional level, from non-NAEP sources: By urban district, school year 2000–01

	Number of schools with students [2]	Number of schools by instructional level [1]			
		Primary	Middle	High	Other
National	90,711	52,285	15,584	17,280	5,562
Central city	21,900	13,962	3,318	3,081	1,539
Atlanta	98	70	15	12	1
Chicago	596	472	21	72	31
District of Columbia	165	112	10	16	27
Houston	289	197	43	35	14
Los Angeles	659	456	75	105	23
New York City	1,207	692	210	184	121

[1] Instructional level is based on the lowest and highest grade in a school: Primary schools begin between prekindergarten and grade 3 and may go as high as grade 8. Middle schools have grade spans ranging from as low as grade 4 to as high as grade 9. High schools start at grade 7 or higher and must extend through grade 12. Other schools include all other grade combinations, including prekindergarten, kindergarten, or 1–12, and ungraded schools.
[2] Includes only schools for which student membership was reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

BEST COPY AVAILABLE

Table B.7 Median pupil/teacher ratios in public elementary and secondary schools, by instructional level, from non-NAEP sources: By urban district, school year 2000–01

	Median pupil/teacher ratio[1]				
		By school instructional level[2]			
	Overall	Primary	Middle	High	Other
National	15.9	16.3	15.7	15.1	12.1
Central city	16.6	16.7	16.0	17.1	11.2
Atlanta	14.2	13.7	16.2	16.2	13.2
Chicago	18.4	18.9	17.3	15.5	11.9
District of Columbia	13.5	13.7	13.6	13.6	8.7
Houston	18.5	18.7	18.7	18.0	3.3
Los Angeles	19.6	19.2	23.6	22.7	10.4
New York City	16.1	15.7	16.7	18.7	6.3

[1] Includes only those schools whose student membership was greater than zero.
[2] Instructional level is based on the lowest and highest grade offered in a school. Primary schools begin between prekindergarten and grade 3 and may go as high as grade 8. Middle schools have grade spans ranging from as low as grade 4 to as high as grade 9. High schools start at grade 7 or higher and must extend through grade 12. Other schools include all other grade combinations, including prekindergarten, kindergarten, or 1–12, and ungraded schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

BEST COPY AVAILABLE

Table B.8 Percentage of staff in public elementary and secondary schools, by type of staff, from non-NAEP sources: By urban district, school year 2000–01

		Percent of full-time equivalent staff						
	Total staff	Teachers	Instructional support	Guidance counselors	Library media staff	District administrators	School administrators	Other staff
National	5,397,788	52.6	11.4	1.8	1.5	1.0	2.5	29.1
Central city	1,513,730	53.4	11.2	1.7	1.3	0.8	2.5	29.1
Atlanta	7,552	52.3	12.7	2.0	1.3	0.5	2.5	28.8
Chicago	28,687 [1]	82.4	2.9	2.9	1.7	1.7	3.1	4.4
District of Columbia	10,808	46.7	10.8	1.9	1.3	0.1	2.5	36.8
Houston	24,820	45.1	10.2	1.2	0.9	0.1	2.3	40.2
Los Angeles	66,598	52.8	14.5	1.4	0.1	0.6	1.9	28.7
New York City	100,198	65.1	0.8	1.9	0.7	0.7	2.9	27.9

[1] The non-teaching staff categories may be underrepresented.
NOTE: Percentage may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Local Education Agency Universe Survey, 2000–01, version 1a.

Table B.9 Percentage of reported students eligible for free/reduced-price school lunch and percentage of students in each racial/ethnic category, from non-NAEP sources: By urban district, school year 2000–01

		Students eligible for free/reduced-price lunch[1]		Racial/ethnic composition of district as a percentage of students enrolled				
	Number of schools with students[2]	Percentage of students	Percentage of schools reporting	White	Black	Hispanic	Asian/ Pacific Islander	American Indian/ Alaska Native
National	90,711	38.6	83.8	61.0	17.0	16.6	4.2	1.2
Central city	21,900	53.6	82.9	36.8	29.2	27.7	5.5	0.9
Atlanta	98	76.4	100.0	6.8	89.5	2.8	0.9	0.1
Chicago	596	—	0.0	9.6	52.0	34.9	3.3	0.2
District of Columbia	165	70.0	98.8	4.5	84.6	9.2	1.6	0.1
Houston	289	70.7	99.7	10.0	32.1	55.0	2.9	0.1
Los Angeles	659	73.5	99.8	9.9	12.8	70.8	6.3	0.3
New York City	1,207	71.9	100.0	15.3	34.9	37.8	11.7	0.3

— Not available.

[1] These percentages should be interpreted with caution; jurisdictions may not have reported students eligible for reduced-price meals, and a number of jurisdictions reported participation instead of eligibility data, which may not be strictly comparable. Percentages are based on those schools that reported.

[2] Includes only schools for which student membership was reported.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School Universe Survey, 2000–01, version 1a, and Local Education Agency Universe Survey, 2000–01, version 1a.

BEST COPY AVAILABLE

Appendix C

Overview of Procedures Used for the NAEP 2002 Writing Trial Urban District Assessment

This appendix provides an overview of the NAEP 2002 writing assessment's primary components—framework, development, administration, scoring, and analysis. A more extensive review of the procedures and methods used in the writing assessment can be found in the forthcoming NAEP 2002 technical report.

The NAEP 2002 Writing Assessment

The NAEP 2002 writing assessment is based on the 1998 writing assessment framework, whose purpose is to provide, based on the expert opinions of writing educators and researchers, a definition of writing to guide the NAEP writing assessment.[1] The framework development process involved the critical input of hundreds of individuals across the country, including representatives of national education organizations, teachers, parents, policymakers, business leaders, and the interested general public. The process was managed by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) under the direction of the National Assessment Governing Board (NAGB); the exercise specifications were developed under contract by American College Testing (ACT) under the direction of NAGB.

[1] National Assessment Governing Board. (2002). *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress.* Washington, DC: Author.

The writing framework delineates six major objectives to organize the design of the assessment.

- Students should write for three major purposes: narrative, informative, and persuasive. While other types of writing could have been included, the developers of the framework believed that, for the purpose of monitoring student achievement (as opposed to creating individual diagnostic assessments), three broad types of writing are appropriate.
- Students should be able to perform a variety of writing tasks (letters, essays, stories, reports) addressing different audiences (peers, school or government officials, business representatives).
- Student writing should be prompted by a variety of stimulus materials (letters, poems, graphics, reports) under varying time constraints.
- Because writing is a dynamic process through which the writer constructs meaning, students should develop their own writing processes, including methods for drafting, evaluating, revising, and editing ideas and forms of expression. Students are to be given planning space in the test materials to generate ideas for drafts. In addition, they are given a pamphlet with suggestions for planning, revising, and editing. All NAEP student responses, given assessment time constraints, are to be evaluated as first drafts.
- Students should display effective choices in the organization of their writing. Further, they should be able to illustrate and elaborate their ideas and should use appropriate conventions of English. All of these characteristics are to be part of the evaluation of student writing.
- Students should value writing as a communicative activity.

Figure C.1 gives examples of various writing tasks similar to those included in the assessment at grades 4 and 8. Included in the figure are descriptions of sample tasks that illustrate how each purpose for writing is assessed.

Figure C.1 Illustrative examples of writing tasks, by purpose for writing, grades 4 and 8

Purpose for writing	Grade 4	Grade 8
Narrative	Provide visual stimuli of a season of the year. Ask students to write a letter to a grandparent telling the story of an interesting personal experience related to the season.	Provide visual stimuli. Ask students to write an article for a sports magazine telling the story of a time when they participated in a hobby or skill they enjoyed.
Informative	Provide an appropriate quotation. Ask students to explain in an essay to their English teacher how a person (parent, teacher, friend) has influenced them in an important way.	Provide a series of brief journal entries from another historical time. Ask students to explain what is revealed about the person who wrote the entries.
Persuasive	Provide visual stimuli of an animal. Ask students to convince their parents/guardians of an animal that would make the best pet.	Provide brief reviews, as models, of a film, TV program, or book. Ask students to write a review for the school newspaper that will convince other students to watch a favorite film or TV program or read a favorite book.

SOURCE: National Assessment Governing Board. *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress.* Washington, DC: Author.

In addition to the six objectives, the framework specifies the percentage of the writing tasks for student responses in the assessment that should be devoted to each of the three writing purposes—narrative, informative, and persuasive. The actual percentage distributions of writing tasks in the assessment are listed in table 1.1 of chapter 1. The table shows the number of tasks at each grade level for each purpose. These target percentages vary by grade level according to what is deemed developmentally appropriate for each grade, as stated in the writing framework.

The Assessment Design

Each student who participated in the writing assessment received a booklet containing two 25-minute writing tasks. All student responses to the writing tasks were rated according to a six-level scoring guide. In addition, the test booklets contained general background questions and writing-specific background questions.

The assessment design allowed for maximum coverage of the writing domain at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of tasks, in which each student was given only 2 of the 20 or more tasks at each grade level. Representative samples of students responded to each task, so that the aggregate results across the entire assessment allow broad reporting of writing abilities for the targeted population.

In addition to matrix sampling, the assessment design used a procedure for distributing blocks across booklets that controlled for position and context effects. Students received different blocks of tasks in their booklets according to a procedure called partially balanced incomplete block (pBIB) spiraling. The procedure assigned blocks of questions in a manner that balanced the positioning of blocks across booklets and balanced the pairing of blocks within booklets according to purposes for writing. Blocks were balanced within each purpose for writing and were partially balanced across purposes for writing. (The spiraling aspect of this procedure cycles the booklets for administration so that, typically, only a few students in any assessment session receive the same booklets.)

In addition to the student assessment booklets, three other instruments provided data relating to the assessment—a teacher questionnaire, a school questionnaire, and a questionnaire regarding students with disabilities and limited English proficient students (SD/LEP). The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students who were selected to participate in the assessment and who were identified as having an Individualized Education Program (IEP) or equivalent plan and/or being limited English proficient (LEP). An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire asked about the student and the special programs in which he or she participated.

NAEP Samples

National Sample

The national results presented in the *Nation's Report Card: Writing 2002* are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. At grades 4 and 8, the national sample was a subset of the combined sample of students assessed in each participating state and the District of Columbia, plus an additional sample from the states that did not participate in the state assessment, and a private school sample. In accordance with the NAEP legislation, the program uses a random selection process in order to obtain a representative sample of students for reporting national and state or jurisdiction results.

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences between the student samples and the populations from which they were drawn. Sampling weights applied to the national and state samples compensate for the disproportionate representation due to oversampling of students who attend nonpublic schools and schools with high concentrations of Black and/or Hispanic students. Among other uses, sampling weights also compensate for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.[2] Appropriate sampling weights were applied to the Trial Urban District Assessment samples.

[2] Additional details regarding the design and structure of the national and state samples may be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

Testing accommodations (e.g., extended time, small group testing) were permitted for special-needs students selected to participate in the NAEP writing assessments. NAEP inclusion rules were applied, and accommodations were offered when a student had an Individualized Education Program (IEP) because of a disability, was protected under Section 504 of the Rehabilitation Act of 1973[3] because of disability, and/or was identified as being a limited English proficient student (LEP); all other students were asked to participate in the assessment under standard conditions.

Standards for Sample Participation and Reporting of Results

In carrying out the NAEP 2002 state assessments and Trial Urban District Assessment, the National Center for Education Statistics (NCES) established participation rate standards that jurisdictions were required to meet in order for their results to be reported. NCES also established additional standards that required the annotation of published results for jurisdictions whose sample participation rates were low enough to raise concerns about their representativeness. The NCES guidelines used to report results in the state assessments and Trial Urban District Assessment, and the guidelines for notation when there is some risk of nonresponse bias in the reported results, are presented in this section.

Guideline 1

The publication of NAEP results

The conditions that will result in the publication of a jurisdiction's results are presented below.

Guideline 1 — Publication of Public School Results

A jurisdiction will have its public school results published in the NAEP 2002 writing report card (or in other reports that include all state-level results) if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent. Similarly, a jurisdiction will receive a separate NAEP state report if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent.

Discussion: If a jurisdiction's public school participation rate for the initial sample of schools is below 70 percent, there is a substantial possibility that bias will be introduced into the assessment results. This possibility remains even after making statistical adjustments to compensate for school nonparticipation. There remains the likelihood that, in aggregate, the substitute schools are sufficiently dissimilar from the originals they are replacing and represent too great a proportion of the population to discount such a difference. Similarly, the assumptions underlying the use of statistical adjustments to compensate for nonparticipation are likely to be significantly violated if the initial response rate falls below the 70 percent level. Guideline 1 takes this into consideration. This guideline is congruent with current NAGB policy, which requires that data for jurisdictions that do not have a 70 percent before-substitution participation rate be reported "in a different format," and with the Education Information Advisory Committee (EIAC) resolution, which calls for data from such jurisdictions not to be published.

[3] Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

The guidelines concerning school and student participation rates in the NAEP state and Trial Urban District Assessments were established to address four significant ways in which nonresponse bias could be introduced into the jurisdiction sample estimates: overall school nonresponse, strata-specific school nonresponse, overall student nonresponse, and strata-specific student nonresponse. Presented on the following pages are the conditions that will result in a jurisdiction or district's receiving a notation in the 2002 reports. Note that in order for a jurisdiction/district's results to be published with no notations, that jurisdiction or district must satisfy all guidelines.

Guideline 2

Reporting school and student participation rates with possible bias due to school nonresponse

Guideline 2—Notation for Overall Public School Participation Rate

A jurisdiction that meets Guideline 1 will receive a notation if its weighted participation rate for the initial sample of public schools was below 85 percent and the weighted public school participation rate after substitution was below 90 percent.

Discussion: For jurisdictions that did not use substitute schools, the participation rates are based on participating schools from the original sample. In these situations, the NCES standards specify weighted school participation rates of at least 85 percent to guard against potential bias due to school nonresponse. Thus the first part of these guidelines, referring to the weighted school participation rate for the initial sample of schools, is in direct accordance with NCES standards.

To help ensure adequate sample representation for each jurisdiction participating in the NAEP 2002 state assessments, NAEP provided substitutes for nonparticipating public schools. For jurisdictions that used substitute schools, the assessment results will be based on the student data from all schools participating from both the original sample and the list of substitutes (unless both an initial school and its substitute eventually participated, in which case only the data from the initial school will be used).

The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that decide not to participate in the assessment. However, considerable technical consideration was given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate bias due to the nonparticipation of initially selected schools. Thus, for the weighted school participation rates including substitute schools, the guidelines were set at 90 percent. (NOTE: *There was no substitution of schools for the Trial Urban District Assessment*)

If a jurisdiction meets either standard (i.e., 85 percent or higher prior to substitution or 90 percent or higher after substitution), there will be no notation for the relevant overall school participation rate.

Guideline 3

Important segments of the jurisdiction's student population that must be adequately represented to avoid possible nonresponse bias

Guideline 3 — Notation for Strata-Specific Public School Participation Rates

A jurisdiction that is not already receiving a notation under Guideline 2 will receive a notation if the sample of public schools included a class of schools with similar characteristics that had a weighted participation rate (after substitution) of below 80 percent, and from which the nonparticipating schools together accounted for more than 5 percent of the jurisdiction's total weighted sample of public schools. The classes of schools from each of which a jurisdiction needed minimum school participation levels were determined by degree of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, if some important segment of the jurisdiction's population is not adequately represented, it is of concern, regardless of the overall participation rate.

If nonparticipating schools are concentrated within a particular class of schools, the potential for substantial bias remains, even if the overall level of school participation appears to be satisfactory. Nonresponse adjustment cells for public schools have been formed within each jurisdiction, and the schools within each cell are similar with respect to degree of urbanization, minority enrollment, and/or median household income, as appropriate for each jurisdiction.

If the weighted response rate, after substitution, for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the sampled schools are nonparticipants from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific school response rates.

Guideline 4

Possible student nonresponse bias

Guideline 4 — Notation for Overall Student Participation Rate in Public Schools

A jurisdiction that meets Guideline 1 will receive a notation if the weighted student response rate within participating public schools was below 85 percent.

Discussion: This guideline follows the NCES standard of 85 percent for overall student participation rates. The weighted student participation rate is based on all eligible students from initially selected or substitute schools who participated in the assessment in either an initial session or a make-up session. If the rate falls below 85 percent, the potential for bias due to students' nonresponse is too great.

Guideline 5

Possible nonresponse bias from inadequately represented strata

Guideline 5 — Notation for Strata-Specific Student Participation Rates in Public Schools

A jurisdiction that is not already receiving a notation under Guideline 4 will receive a notation if the sampled students within participating public schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than 5 percent of the jurisdiction's weighted assessable public school student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the student, whether or not the student was classified as a student with a disability (SD) or limited English proficient (LEP), and the type of assessment session, as well as school level of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: This guideline addresses the fact that if nonparticipating students are concentrated within a particular class of students, the potential for substantial bias remains, even if the overall student participation level appears to be satisfactory. Student nonresponse adjustment cells have been formed using the school-level nonresponse adjustment cells, together with the student's age and the nature of the assessment session.

If the weighted response rate for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the invited students who do not participate in the assessment are from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific student response rates.

In the 2002 Trial Urban District Assessment, New York City did not meet the initial public-school participation rate standard of 70 percent at the eighth grade. Consequently the performance results for this grade were not reported.

At grade 4, New York City also did not meet the second guideline (i.e., the weighted participation rate for the initial sample of schools was below 85 percent and the weighted school participation rate after substitution was below 90 percent).

Results for New York City at grade 4 are shown with a notation indicating possible bias related to nonresponse.

Trial Urban District Assessment Samples

Sampling for the Trial Urban District Assessment was modeled on NAEP's state sampling procedures. However, school substitution, which is used in state assessments, was not an option in the Trial Urban District Assessment. Tables C.1 and C.2 provide a summary of the 2002 urban district school and student participation rates for the writing assessment sample. The first rate in each table is the weighted percentage of schools participating in the assessment. This rate is based on the number of schools that were selected for the assessment. The numerator of this rate is the estimated number of students represented by each selected school that participated in the assessment. The denominator is the estimated number of students represented by each of the selected schools that had eligible students enrolled.

Also presented in tables C.1 and C.2 are weighted student participation rates. The numerator of this rate is the estimated number of students that each student represents across all students *assessed* (in either an initial session or a makeup session). The denominator of this rate is the estimated number of students represented in the sample, across all *eligible* sampled students in participating schools. The number of students that each student represents is mainly determined by the probability that a student is included in the sample, with necessary adjustments made for other factors. The overall participation rates take into account the weighted percentage of student participation after makeup sessions.

BEST COPY AVAILABLE

Table C.1 Weighted school and student participation rates, grade 4: By urban district, 2002

Grade 4	Number of schools	Number of students	School rate	Student rate	Overall rate
Atlanta	49	1,501	98	93	92
Chicago	76	2,037	95	92	87
District of Columbia	117	2,750	100	90	90
Houston	49	1,321	98	95	93
Los Angeles	76	2,005	100	93	93
New York City [1]	38	924	76	89	67

[1] Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

Table C.2 Weighted school and student participation rates, grade 8: By urban district, 2002

Grade 8	Number of schools	Number of students	School rate	Student rate	Overall rate
Atlanta	15	1,296	100	91	91
Chicago	69	1,611	94	92	86
District of Columbia	36	1,856	100	85	85
Houston	34	1,109	96	89	86
Los Angeles	66	1,778	97	90	87
New York City [1]	31	683	63	81	51

[1] Indicates that the district did not meet the guideline for 70 percent school participation in 2002.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

Results from the 2002 writing assessments are reported (on a trial basis) here on district-level samples of fourth- and eighth-grade students in the large urban school districts that participated in the Trial Urban District Assessment. Results for fourth- and eighth-graders in the District of Columbia, which has been assessed in the past as a jurisdiction, are also reported. The samples of students in the urban school districts represent augmentations of students who would normally be selected as part of state samples. These samples allow reliable subgroup reporting in these districts. All students at "lower" geographical levels are assumed to be part of "higher-level" samples. For example, Houston is one of the urban districts included in the Trial Urban District Assessment. Data from students tested in the Houston sample are used to report results for Houston and also contribute to the Texas estimates and to the national calculations.

Tables C.3 and C.4 display the target student and school sample sizes planned for the Trial Urban District Assessment. The first column contains the planned number of schools for each district. The second column contains the number of schools that would have been sampled by NAEP in each district had there been no Trial Urban District Assessment. The last column shows the planned student sample size. Note that the District of Columbia is not presented in the tables because its normal sampling plan did not have to be adjusted for the Trial Urban District Assessment.

Table C.3 Number of schools and students planned for the Trial Urban District Assessment, grade 4: By urban district, 2002

Grade 4	Number of schools for trial assessment	Number of schools normally sampled in NAEP	Number of students planned for trial assessment
Atlanta	51	6	1250
Chicago	81	24	1900
Houston	51	5	1250
Los Angeles	76	12	1900
New York City [1]	52	41	1250

[1] Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

Table C.4 Number of schools and students planned for the Trial Urban District Assessment, grade 8: By urban district, 2002

Grade 8	Number of schools for trial assessment	Number of schools normally sampled in NAEP	Number of students planned for trial assessment
Atlanta	17	8	1250
Chicago	79	26	1900
Houston	39	6	1250
Los Angeles	68	13	1900
New York City [1]	52	37	1250

[1] Indicates that the district did not meet the guideline for 70 percent school participation in 2002.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

Note that the sample sizes vary among districts. The study was designed to allow the examination of the quality of data that resulted from samples of different sizes. Therefore, larger samples were selected in two of the districts (Chicago and Los Angeles) and smaller samples were selected in the remaining three.

Data Collection and Scoring

The NAEP 2002 writing assessment was conducted from January to March 2002. Data collection for the NAEP 2002 assessment at the national, state, and district levels was conducted by trained field staff from Westat.

Materials from the NAEP 2002 assessment were shipped to Pearson, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by Educational Testing Service (ETS). All the writing tasks were evaluated according to six-level scoring guides. At each grade, scoring guides were developed for each of the three types of writing purpose: narrative, informative, and persuasive.

Specialists in writing who are highly experienced in teaching and/or assessing writing trained the professional raters who evaluated the student responses. The trainers received intensive training together, which included reading a manual that explained how to use the scoring guides and the processes for training and checking raters. For each task, or writing problem, the trainer, in consultation with other trainers or assessment specialists, chose numerous sample responses to present to raters and prepared notes on how the scoring guide applied. The sample responses helped raters become accustomed to the variety of responses the task elicited before they began rating the student responses. Raters had to pass a qualifying test before they could evaluate student responses: they had to agree with at least 70 percent of the ratings (to a set of ten student responses) that were given beforehand by their trainer. In order to determine interrater reliability of scoring, a specified percentage of responses was read twice: 6 percent of the responses at grades 4 and 8 were read by two raters.

In 2002, raters scored a total of 608,269 student responses to writing tasks for the national, state, and district assessments. This number includes rescoring to monitor interrater reliability. The average percentages of exact agreement of ratings on the six-level scoring guides for the 2002 reliability samples were 83 percent at the fourth grade and 82 percent at the eighth grade.

Data Analysis and Item Response Theory (IRT) Scaling

After the professional scoring, all information was transcribed into the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information was compiled in the database, the data were weighted according to the population structure. The weighting for the national, state, and urban district samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.[4]

Analyses were then conducted to determine the percentages of students whose responses to each writing task

[4] Weighting procedures are described more fully in the "Weighting and Variance Estimation" section in this document. Additional information about the use of weighting procedures may be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

BEST COPY AVAILABLE

attained each level on the scoring guide, and who provided various responses to each background question. In calculating response percentages for each task, only students classified as having been presented with the question were included in the denominator of the statistic. Students whose papers were blank or whose responses were judged to be off-topic were excluded from the calculation of the scale.

Item Response Theory (IRT) was used to estimate average writing scale scores for the nation, for various subgroups of interest within the nation, for the states, and for other jurisdictions. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. NAEP used IRT analysis to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity.

The results for the NAEP 2002 writing assessments are presented on the NAEP writing scales. In 2002, a scale ranging from 0 to 300 was created to report performance at each grade level. The scale summarizes student performance across all three purposes for writing (narrative, informative, and persuasive) in the assessment.

In producing the writing scale, an IRT model was used. The writing tasks (all rated according to six-level scoring guides) were scaled by use of a Generalized Partial-Credit (GPC) model.[5] Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories that are used for more complex constructed-response questions such as writing tasks.[6]

Because of the assessment booklet design used by NAEP, students do not receive enough writing tasks to provide reliable information about individual performance. Traditional test scores for individual students, even those based on the IRT, would result in misleading estimates of population characteristics, such as subgroup means and percentages of students performing at or above a certain scale score level. However, it is NAEP's goal to estimate these population characteristics. NAEP's objectives can be achieved with methodologies that produce estimates of the population-level parameters directly, without the intermediary computation of estimates of individuals.[7] This is accomplished using marginal estimation scaling model techniques for latent variables. Under the assumptions of the scaling models, these population estimates will be consistent in the sense that the estimates approach the model-based population values as the sample size increases. This would not be the case for population

[5] Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

[6] More detailed information regarding the IRT analyses used in NAEP may be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

[7] Mislevy, R. J., and Sheehan, K. M. (1987). Marginal Estimation Procedures. In A. E. Beaton (Ed.) *Implementing the New Design: The NAEP 1983–1984 Technical Report* (Rep. No. 15-TR-20), pp. 293–260. Princeton, NJ: Educational Testing Service.

estimates obtained by aggregating optimal estimates of individual performance.[8]

Weighting and Variance Estimation

A complex sampling design was used to select the students who were assessed. The properties of a sample selected through such a design may be very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data were estimated using sampling weights. These weights included adjustments for school and student nonresponse.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student ability: 1) the uncertainty due to sampling only a relatively small number of students, and 2) the uncertainty due to sampling only a portion of the cognitive domain of interest. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who had a certain rating for their responses to a task.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only two writing tasks, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of groups and subgroups of students. The estimate of the variance of the students' scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.[9]

---

[8] For theoretical and empirical justification of the procedures employed, see Mislevy, R. J. (1988). Randomization-Based Inferences About Latent Variables from Complex Samples. *Psychometrika 56*(2), 177–196. Donoghue, J. R. (1994). An Empirical Examination of the IRT Information of Polytomously Scored Writing Items Under the Generalized Partial Credit Model. *Journal of Educational Measurement 31*(4), 295–311.

[9] For further details, see Johnson, E. G., and Rust, K. F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics 17*(2), 175–190.

Typically, when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the estimation of standard errors may be quite large. Additional details concerning procedures for identifying such standard errors will be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources—inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate and, because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

Drawing Inferences from the Results

The reported statistics are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty related to the fact that they cannot ask all questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the estimated standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account, and observed similarities or differences should not be relied on solely. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of those statistics and the magnitude of the difference among the averages or percentages.

For the data presented in this report, all the estimates have corresponding estimated standard errors. For example, table C.5 shows the average scale score for the NAEP 2002 Trial Urban District Assessments and percentages of students at or above proficiency levels by gender for grade 4. Table C.6 shows the scores and standard errors for the 25th, 50th, and 75th percentiles at grades 4 and 8. Estimated standard errors appear in parentheses next to each estimated scale score or percentage. For the estimated standard errors corresponding to other data in this report, the reader can consult the NCES web site at http://nces.ed.gov/nationsreportcard/naepdata/.

Table C.5 Average writing scale scores, percentage of students at or above each achievement level, and standard errors, by gender, grade 4 public schools: By urban district, 2002

Grade 4	Percentage of students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	*Advanced*
Male						
Nation (Public)	51 (0.3)	144 (0.6)	20 (0.5)	80 (0.5)	18 (0.4)	1 (0.1)
Central city (Public) [1]	51 (0.4)	139 (0.7) **	24 (0.9) **	76 (0.9) **	15 (0.7) **	1 (0.1)
Atlanta	47 (1.1)	131 (2.2) *,**	31 (3.1) *,**	69 (3.1) *,**	8 (1.6) *,**	1 (0.4)
Chicago	50 (1.1)	130 (1.9) *,**	32 (2.6) *,**	68 (2.6) *,**	7 (1.2) *,**	# (***)
District of Columbia	49 (1.1)	127 (1.1) *,**	37 (1.5) *,**	63 (1.5) *,**	7 (0.9) *,**	# (0.2) *
Houston	48 (1.7)	141 (2.8)	24 (3.2)	76 (3.2)	17 (2.6)	1 (0.5)
Los Angeles	51 (1.1)	134 (2.1) *,**	29 (2.6) **	71 (2.6) **	11 (1.7) *,**	# (***)
New York City [1]	50 (2.0)	145 (4.1)	21 (3.2)	79 (3.2)	21 (4.1)	1 (0.6)
Female						
Nation (Public)	49 (0.3)	162 (0.4)	10 (0.4)	90 (0.4)	35 (0.6)	3 (0.2)
Central city (Public) [1]	49 (0.4)	155 (0.8) **	13 (0.6) **	87 (0.6) **	27 (0.9) **	2 (0.3) **
Atlanta	53 (1.1)	147 (2.7) *,**	17 (2.9) **	83 (2.9) **	18 (2.8) *,**	2 (0.8)
Chicago	50 (1.1)	145 (2.0) *,**	17 (2.0) **	83 (2.0) **	16 (2.6) *,**	# (***)
District of Columbia	51 (1.1)	143 (1.4) *,**	18 (1.5) *,**	82 (1.5) *,**	15 (1.5) *,**	1 (0.3) *,**
Houston	52 (1.7)	154 (3.3) **	15 (2.4) **	85 (2.4) **	28 (3.2)	3 (0.9)
Los Angeles	49 (1.1)	148 (2.5) *,**	16 (2.3) **	84 (2.3) **	21 (2.6) *,**	1 (0.4) *,**
New York City [1]	50 (2.0)	160 (2.9)	10 (3.0)	90 (3.0)	33 (4.0)	3 (1.1)

Percentage rounds to zero.
[1] Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.
* Significantly different from central city public schools.
** Significantly different from nation (public schools).
(***) Standard error estimates cannot be accurately determined.
[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.
NOTE: Standard errors of the estimated percentages and scale scores appear in parentheses. Percentages below and at or above *Basic* may not add to 100, due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

BEST COPY AVAILABLE

Table C.6 Selected writing percentiles and estimated standard errors, grade 4 and 8 public schools: By urban district, 2002

	Scale score distribution		
	25th percentile	50th percentile	75th percentile
Grade 4			
Nation (Public)	128 (0.6)	153 (0.5)	178 (0.4)
Central city (Public) [1]	122 (0.7) **	146 (0.7) **	171 (0.9) **
Atlanta	117 (1.9) *,**	139 (2.2) *,**	161 (2.1) *,**
Chicago	116 (2.6) *,**	137 (1.4) *,**	160 (2.1) *,**
District of Columbia	113 (1.1) *,**	134 (1.8) *,**	157 (1.6) *,**
Houston	123 (3.7)	147 (3.0)	174 (4.4)
Los Angeles	117 (3.0) **	141 (1.8) *,**	165 (2.8) *,**
New York City [1]	128 (3.8)	154 (2.5) *	178 (4.2)
Grade 8			
Nation (Public)	127 (0.7)	153 (0.6)	178 (0.7)
Central city (Public) [2]	117 (0.8) **	144 (1.1) **	170 (1.4) **
Atlanta	107 (2.0) *,**	129 (2.2) *,**	151 (1.6) *,**
Chicago	111 (3.4) **	136 (2.2) *,**	160 (3.9) *,**
District of Columbia	105 (1.7) *,**	128 (1.2) *,**	152 (2.0) *,**
Houston	113 (4.3) *,**	139 (2.3) **	165 (3.1) **
Los Angeles	104 (1.7) *,**	128 (1.5) *,**	152 (2.0) *,**

[1] Although deemed sufficient for reporting, the target response rate specified in the NAEP guidelines was not met.

* Significantly different from central city public schools.

** Significantly different from nation (public schools).

[1] For comparison, at fourth grade 66 percent of students in central city public schools and 40 percent in public schools nationally were non-White. Also, 61 percent of students in central city public schools and 43 percent in public schools nationally were eligible for free/reduced-price lunch.

[2] For comparison, at eighth grade 61 percent of students in central city public schools and 36 percent in public schools nationally were non-White. Also, 48 percent of students in central city public schools and 34 percent in public schools nationally were eligible for free/reduced-price lunch.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Trial Urban District Writing Assessment and 2002 Writing Assessment.

Using confidence intervals based on the standard errors provides a way to take into account the uncertainty associated with sample estimates and to make inferences about the population averages and percentages in a manner that reflects that uncertainty. An estimated sample average scale score plus or minus 1.96 standard errors approximates a 95 percent confidence interval for the corresponding population quantity. This statement means that one can conclude with an approximately 95 percent level of confidence that the average performance of the entire population of

BEST COPY AVAILABLE

interest (e.g., all fourth-grade students in public schools) is within plus or minus 1.96 standard errors of the sample average.

For example, suppose that the average writing scale score of the students in a particular group was 156 with an estimated standard error of 1.2. An approximately 95 percent confidence interval for the population quantity would be as follows:

Average [plus or minus] 1.96 standard errors

$$156 \pm 1.96 \times 1.2$$

$$156 \pm 2.4$$

(153.6, 158.4)

Thus, one can conclude with a 95 percent level of confidence that the average scale score for the entire population of students in that group is between 153.6 and 158.4. It should be noted that this example and the examples in the following sections are illustrative. More precise estimates carried out to one or more decimal places are used in the actual analyses.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. Extreme percentages should be interpreted with caution. Adding or subtracting the standard errors associated with extreme percentages could cause the confidence interval to exceed 100 percent or fall below 0 percent, resulting in numbers that are not meaningful. A more complete discussion of extreme percentages may be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether the evidence based on the data from the groups in the sample is strong enough to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the "standard error of the difference" between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

Standard Error of the Difference =

$$SE_{A-B} = \sqrt{(SE_A^2 + SE_B^2)}$$

The standard error of the difference, like the standard error for an individual group average or percentage, can be used to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the 0.05 level.

The following example of comparing groups addresses the problem of determining whether the average writing scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are as follows:

Group	Average Scale Score	Standard Error
A	137	0.9
B	135	1.1

The difference between the estimates of the average scale scores of groups A and B is two points (137–135). The estimated standard error of this difference is

$$\sqrt{(0.9^2 + 1.1^2)} = 1.4$$

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference.

$$2 \pm 1.96 \times 1.4$$

$$2 \pm 2.7$$

$$(-0.7, 4.7)$$

The value zero is within the confidence interval; therefore, there is insufficient evidence to claim that group A outperformed group B.

Conducting Multiple Tests

The procedures in the previous section and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, there are times when many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), adjustments (called "multiple comparison procedures")[10] must be made to the methods described in the previous section. One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure was used to control the certainty level.[11]

[10] Miller, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlag.

[11] Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, Series B, no. 1, 289–300.

Unlike the other multiple comparison procedures that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.[12] Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures. A detailed description of the FDR procedure may be found in the technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

To illustrate how the FDR procedure is used, consider the comparisons of male and female average writing scale scores for the five groups presented in table C.7. Note that the difference in average scale scores and the estimated standard error of the difference are calculated in a way comparable with that of the example in the previous section. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference.

Table C.7 Example of FDR comparisons of average scale scores for different groups of students

	Male		Female		Male and female			
	Average scale score	Standard error	Average scale score	Standard error	Difference in averages	Standard error of difference	Test statistic	Percent confidence[1]
Group 1	224	1.3	226	1.0	2.08	1.62	1.29	20
Group 2	187	1.7	193	1.7	6.31	2.36	2.68	1
Group 3	191	2.6	197	1.7	6.63	3.08	2.15	4
Group 4	229	4.4	232	4.6	3.24	6.35	0.51	62
Group 5	201	3.4	196	4.7	-5.51	5.81	-0.95	35

[1] The percent confidence is 2(1-F(x)) where F(x) is the cumulative distribution of the t-distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.

The difference in average scale scores and the estimated standard error of that difference can be used to find an approximately 95 percent confidence interval, as in the example in the previous section, or they can be used to identify a confidence percentage. In the example in the previous section, because an approximately 95 percent confidence interval was desired, the number 1.96 was used to multiply the estimated standard error of the difference to create the approximate confidence interval. In the current example, the confidence interval for the test statistics is identified from statistical tables. Instead of checking to see if zero is within the 95 percent confidence interval about the mean, the significance level from the statistical tables can be directly compared to 100 − 95 = 5 percent.

If the comparison of average scale scores for males and females were made for any one of the five groups, there would be a significant difference between the average

[12] Williams, V. S. L., Jones, L. V., and Tukey, J. W. (1999). Controlling Error in Multiple Comparisons with Examples From State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics 24*(1), 42–69.

scale scores for the two gender groups at the significance level of less than 5 percent. However, because we are interested in the difference in average scale scores across genders for all five of the groups, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons. However, comparisons of average scale scores for each male/female pair were treated separately, so the steps described in this example would be replicated for the comparison of other male and female average scale scores.

Using the FDR procedure to take into account that all comparisons are of interest to us, the percentages of confidence in the example are ordered from largest to smallest: 62, 35, 20, 4, and 1. In the FDR procedure, 62 percent confidence for the group 4 comparison would be compared to 5 percent, 35 percent for the group 5 comparison would be compared to $0.05 \times (5-1)/5 = 0.04 = 4$ percent,[13] 20 percent for the group 1 comparison would be compared to $0.05 \times (5-2)/5 = 0.03 = 3$ percent, 4 percent for the group 3 comparison would be compared to $0.05 \times (5-3)/5 = 0.02 = 2$ percent, and 1 percent for the group 2 comparison (actually slightly smaller than 1 prior to rounding) would be compared to $0.05 \times (5-4)/5 = 0.01 = 1$ percent. The procedure stops with the first contrast found to be significant. If there are any remaining contrasts, they are all declared to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference between the male and female average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for males and females are not significantly different from one another. In practice, a very small number of counterintuitive results occur when the FDR procedures are used to examine differences in subgroup results by jurisdiction. In those cases, results were not included in this report. NCES is continuing to evaluate the use of FDR and multiple-comparison procedures for future reporting.

NAEP Reporting Groups

Results are provided for groups of students defined by shared characteristics—gender, race/ethnicity, school's type of location, and eligibility for free/reduced-price school lunch. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).[14] The first-stage sampling units in the selection of Trial Urban District Assessment samples are schools. However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations are presented below. Note that not all of the reporting groups used for the national report card are included in this report on the urban districts.

[13] The level of confidence times the number of comparisons minus one divided by the number of comparisons is $0.05\ (5-1)/5 = 0.04 = 4$ percent.

[14] For the NAEP national assessments prior to 2002, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). In 2002, the first-stage sampling units are schools (public and nonpublic) in the selection of the combined sample. Further details about the procedure for determining minimum sample size will appear in technical documentation section of the NAEP web site at http://nces.ed.gov/nationsreportcard.

Gender

Results are reported separately for males and females.

Race/Ethnicity

In all NAEP assessments, data about student race/ethnicity is collected from two sources: school records and student self-reports. Previously, NAEP has used student self-reported race as the primary race/ethnicity reporting variable. As of 2002, school-recorded race has become the race/ethnicity variable presented in NAEP reports. The mutually exclusive racial/ethnic categories are White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. When a school reports a student's race as "Other," that category is used. If the school record for race is missing for the student, the student's response to the race/ethnicity question is then used. If student data are missing (i.e. the student did not respond or gave multiple responses), then the student is coded to the "Other/missing" category. The combination of these 2 sets of student categories is used for the "other" category. The race/ethnicity tables in this report omit the "Other" category because the percentages were found to be consistently under one percent. Information based on student self-reported race/ethnicity will continue to be available on the NAEP Data Tool (http://nces.ed.gov/nationsreportcard/naepdata/).

Type of Location

In most NAEP assessments, results are reported for students attending schools in three mutually exclusive location types: central city, urban fringe/large town, and rural/small town. Results for the NAEP 2002 Trial Urban District Assessment are reported for students attending schools in one type of location—central city. Following standard definitions established by the Federal Office of Management and Budget, the U. S. Census Bureau (see http://www.census.gov/) defines "central city" as the largest city of a Metropolitan Statistical Area (MSA) or a Consolidated Metropolitan Statistical Area (CMSA). An MSA is an area defined by the federal government for the purposes of presenting general-purpose statistics for metropolitan areas. Typically, an MSA contains a city with a population of at least 50,000 and includes its adjacent areas. An MSA becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has a population of 1,000,000 or more, its component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation.

In the NCES Common Core of Data (CCD) locale codes are assigned to schools. For the definition of central city used in this report, two locale codes of the survey are combined. The definition of each school's type of location is determined by the size of the place where the school is located and whether or not it is in an MSA or CMSA. School locale codes are assigned by the U.S. Bureau of the Census. For the definition of central city NAEP reporting uses data from two CCD locale codes: large city (a central city of an MSA or CMSA with the city having a population greater than or equal to 250,000) and midsize city (a central city of a MSA or CMSA having a population less than 250,000). Central city is a geographical term and is not synonymous with "inner city."

The boundaries of an urban school district and a city may not always coincide. Los Angeles Unified, for example, extends beyond the city boundaries and includes urban fringe areas of the MSA, although the entire district is coded as central city. Most of the other districts included in this report have school districts that share the same boundaries as the city. The interested reader may view the School District Demographics website at http://nces.ed.gov/surveys/sdds, where the school district's boundaries can be shown on a map that also has county boundaries. In the Houston and Los Angeles districts, some students attended schools located in the category of urban fringe/large town. These included 5 percent and 19 percent of fourth-grade students in Houston and Los Angeles respectively, as well as 24 percent of grade 8 students in Los Angeles. Urban fringe/large town is a NAEP classification that combines three categories: Urban fringe of large city, urban fringe of midsize city, and large town. An urban fringe includes all densely settled places within MSAs that are classified as urban by the U.S. Census Bureau. A large town is defined as a place outside MSAs with a population of less than 25,000 but greater than or equal to 2,500. Across the total sample for the NAEP writing assessment in 2002, 28 percent of students attended schools classified as central city, 42 percent attended schools classified as urban fringe/large town, and 30 percent attended schools classified as rural.

Eligibility for Free/ Reduced-Price School Lunch

Based on available school records, students were classified either as currently eligible for the free/reduced-price school lunch component of the U.S. Department of Agriculture's National School Lunch Program or as not eligible. Eligibility for the program is determined by a student's family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level, and reduced-price lunch qualification is set at 170 percent of the poverty level. The classification applies only to the school year when the assessment was administered (i.e., the 2001–02 school year) and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available."

Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 2002 writing assessment. The assessment was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and Pearson Educational Measurement. Most importantly, NAEP is grateful to the students and school staff who made the assessment possible.

The NAEP 2002 writing assessment was funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. The Associate Commissioners of Education Statistics, Peggy Carr and Val Plisko, and the NCES staff—Arnold Goldstein, Steve Gorman, Carol Johnson, Andrew Kolstad, Taslima Rahman, Marilyn Seastrom, and Holly Spurlock—worked closely and collegially with the authors to produce this report.

The NAEP project at ETS is directed by Stephen Lazer and John Mazzeo, with assistance from John Barone. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities were conducted by Pearson Educational Measurement under the direction of Brad Thayer, Connie Smith, and William Buckles.

Test development activities took place at ETS under the direction of Hilary Persky with assistance from John Morris.

The complex statistical and psychometric activities necessary to report results for the NAEP 2002 reading assessment were directed by Catherine Hombo, Frank Jenkins, and Jiahe Qian, with assistance from Ying Jin.

The extensive data processing and computer programming activities underlying the statistical and psychometric analyses conducted at ETS are under the direction of David Freund, Edward Kulick, Bruce Kaplan, and Steven Isham. Data analyses presented in this report were managed by Scott Davis with assistance from Laura Jerry, Youn-hee Lim, Haiying Liu, Norma Norris, Alfred Rogers, Sarwinder Thind, Mike Weiss, and John Willey. The complex database work for this assessment was managed by Katherine Pashley with assistance from Gerry Kokolis.

The reporting process was directed by Jay Campbell with assistance from Wendy Grigg. The design and production of this report were overseen by Loretta Casalaina with assistance from Joseph Kolodey and Rick Hasney. Ming Kuang coordinated the documentation and data checking procedures with assistance from Janice Goodis, Andrea Bergen, and Alice Kass. Arlene Weiner coordinated the editorial procedures with assistance from Trish Hamill. The Web version of this report was coordinated by Rick Hasney with assistance from Loretta Casalaina.

Many thanks are due to the numerous reviewers external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in the final version of this report: Ellen Carey, James E. Carlson, Kathryn A. Chandler, Young Chun, Mary Crovo, Lawrence Feinberg, Ray Fields, Sharon Lewis, Rita Foy Moss, Sarah Murray, Patrick Rooney, Linda Shafer, Alan Vanneman, and Daisy Vickers.

United States
Department of Education
ED Pubs
8242-B Sandy Court
Jessup, MD 20794-1398

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Official Business Only
Penalty for Private Use, $300

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").